



Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
Biologia Evoluzionistica e Ambientale
Curriculum GENETICA

CICLO XXIII

COORDINATORE Prof. Guido Barbujani

Basi genetiche dell'adattamento alle basse pressioni
di ossigeno in popolazioni di alta quota delle Ande
peruviane.

Settore Scientifico Disciplinare BIO/18

Dottorando

Dott. Chianella Caterina

Tutore

Dott. Fuselli Silvia

Anni 2008/2011

1.INTRODUZIONE	pag. 1
1.1 L'ambiente di alta quota	pag. 1
1.2 Adattamenti fisiologici alla bassa pressione di ossigeno tipica dell'alta quota	pag. 3
1.2.1 Acclimatazione e adattamento	pag. 3
1.2.2 Popolazioni native di alta quota	pag. 6
1.3 Basi genetiche dell'adattamento all'alta quota	pag. 9
1.4 La via metabolica di HIF	pag. 15
1.5 EPO	pag. 19
1.6 VHL	pag. 21
1.7 EPAS1 (HIF-2 α)	pag. 25
2. SCOPO DEL PROGETTO	pag. 26
3. RACCOLTA E DESCRIZIONE DEL CAMPIONE POPOLAZIONISTICO	pag. 27
3.1 I campioni di alta quota	pag. 27
3.1.1 Siti di campionamento	pag. 28
3.1.2 Procedura di campionamento, misurazioni fenotipiche e strumentazione	pag. 31
3.1.3 Risultati delle misurazioni fenotipiche	pag. 35
3.2 Il campione di confronto di bassa quota	pag. 42
4. METODI	pag. 44
4.1 Protocolli di biologia molecolare	pag. 44
4.1.1 Estrazione del DNA	pag. 44
4.1.2 Amplificazioni mediante PCR	pag. 44
4.1.2.1 <i>mtDNA</i>	pag. 44
4.1.2.2 <i>EPO</i>	pag. 46
4.1.2.3 <i>VHL</i>	pag. 48
4.1.2.4 <i>EPAS1</i>	pag. 49
4.1.3 Sequenziamento	pag. 50
4.1.3.1 <i>mtDNA</i>	pag. 50
4.1.3.2 <i>EPO</i>	pag. 51
4.1.3.3 <i>VHL</i>	pag. 51

4.1.3.4 <i>EPAS1</i>	pag. 52
4.2 Analisi statistiche	pag. 54
4.2.1 Determinazione degli aplotipi	pag. 54
4.2.2 Stima del linkage disequilibrium	pag. 56
4.2.3 Variabilità genetica intra-popolazione	pag. 58
4.2.3.1 Indici di diversità	pag. 58
4.2.3.2 Test di neutralità	pag. 59
4.2.4 Variabilità genetica tra popolazioni	pag. 62
4.2.5 Test di associazione	pag. 63
4.2.5.1 Test esatto di Fisher	pag. 63
4.2.5.2 ANOVA (ANalysis Of VAriance)	pag. 64
5. IL CAMPIONE ANDINO NEL CONTESTO SUD-AMERICANO	pag. 66
5.1 Il popolamento dell'America	pag. 66
5.2 Le popolazioni native del Sud America: aspetti genetici	pag. 68
5.3 Il DNA mitocondriale	pag. 70
5.3.1 Distribuzione degli aplogruppi	pag. 72
5.3.2 Variabilità genetica intra-popolazione	pag. 75
5.3.3 Variabilità genetica tra popolazioni	pag. 80
6. IDENTIFICAZIONE ED ANALISI DEI GENI CANDIDATI ALLO STUDIO DELL'ADATTAMENTO GENETICO UMANO ALL'ALTA QUOTA	pag. 81
6.1 Identificazione dei geni candidati	pag. 81
6.1.1 Endeavour	pag. 82
7. EVOLUZIONE MOLECOLARE DEI GENI CANDIDATI	pag. 85
7.1 La teoria neutrale e la stima dei tassi di sostituzione	pag. 86
7.2 PAML: Phylogenetic Analysis by Maximum Likelihood	pag. 90
7.3 Analisi dei geni candidati	pag. 92
7.3.1 Il database di sequenze	pag. 93
7.3.2 CODEML	pag. 93
7.3.3 Prima parte dello studio: <i>Site Models</i>	pag. 96
7.3.4 Seconda parte dello studio: <i>Branch models</i>	pag. 104

7.3.4.1 Il dataset	pag. 103
7.3.4.2 Analisi dell'insieme delle specie adattate all'ipossia	pag. 106
7.3.4.3 Analisi delle singole specie adattate all'ipossia	pag. 107
8. VARIABILITÀ GENETICA DEI GENI CANDIDATI <i>EPO</i> , <i>VHL</i> E <i>EPAS1 (HIF-2α)</i> E ASSOCIAZIONE CON L'ALTA QUOTA	pag. 109
8.1 <i>EPO</i>	pag. 110
8.1.1 <i>EPO</i> : variabilità genetica all'interno e tra popolazioni di alta quota	pag. 111
8.1.2 <i>EPO</i> : confronti alta e bassa quota	pag. 113
8.1.3 <i>EPO</i> : associazione genotipo/fenotipo	pag. 113
8.2 <i>VHL</i>	pag. 116
8.2.1 <i>VHL</i> : variabilità genetica all'interno e tra popolazioni di alta quota	pag. 117
8.2.2 <i>VHL</i> : confronti alta e bassa quota	pag. 120
8.2.3 <i>VHL</i> : associazione genotipo/fenotipo	pag. 121
8.3 <i>EPAS1 (HIF-2α)</i>	pag. 122
9. DISCUSSIONE E CONCLUSIONI	pag. 123
Riferimenti bibliografici	pag. 134
Appendice 1: kit di raccolta saliva (Genotek) e protocollo di estrazione DNA	
Appendice 2: Database completo dei campioni di alta quota	
Appendice 3: Specie utilizzate nei <i>site models</i> e relativi codici identificativi	
Appendice 4: Specie utilizzate nei <i>branch models</i> e relativi codici identificativi	

1. INTRODUZIONE

1.1 L'ambiente di alta quota

Durante gli ultimi 100 mila anni a seguito dell' "*out of Africa*" le popolazioni umane hanno colonizzato ambienti estremamente variabili andando incontro ad un processo di differenziazione che ha permesso l'adattamento a diversi tipi di alimentazione, di climi, di agenti patogeni, come a diversi livelli di altitudine e quindi di ossigeno (Baker 1984, Balaesque et al. 2007, Beall 2007b).

All'aumentare dell'altitudine corrispondono infatti una serie di cambiamenti climatici e atmosferici che hanno un impatto rilevante sull'uomo.

Innanzitutto si osserva una progressiva riduzione della temperatura di circa 1°C ogni 150 metri, indipendentemente dalle condizioni meteorologiche. Questo, combinato alla presenza di vento, frequente in alta quota, riduce il calore prodotto dal corpo umano causando un progressivo raffreddamento. Inoltre si riduce la quantità di vapore acqueo per unità di volume di aria e questo determina una secchezza delle prime vie aeree a cui segue il rischio sempre maggiore di disidratazione per cessione di acqua dal corpo allo scopo di umidificare l'aria inspirata. In alta quota la riduzione dello strato atmosferico che assorbe le radiazioni solari aumenta il rischio di effetti dannosi dei raggi ultravioletti (UV) la cui azione è ulteriormente amplificata dalla presenza di neve o ghiaccio che riflettono i raggi. Allo stesso tempo però una maggiore esposizione dell'ambiente alle radiazioni UV determina una sorta di "sterilizzazione naturale" riducendo la carica batterica e il rischio di infezioni. Questo, combinato a bassi livelli di umidità, riduce anche il rischio di allergie poiché molti aero allergeni, come ad esempio gli acari, necessitano di ambienti umidi per poter sopravvivere (Baker 1984, Cogo 2009).

Tuttavia ciò che più caratterizza l'ambiente di alta quota è la progressiva riduzione della pressione barometrica a cui si accompagna una progressiva diminuzione della pressione parziale di ossigeno (pO_2). La pressione parziale dell'ossigeno (pO_2) nell'atmosfera corrisponde al 20,93% della pressione barometrica, quindi se a livello del mare la pressione barometrica è pari a 760 mmHg e la pressione parziale di ossigeno è di circa

159,2 mmHg, a 3000 m la pressione barometrica scende a 526,3 mmHg e la pressione parziale di ossigeno a 110,2 mmHg.

Questa riduzione di circa il 30% nella pO_2 si riflette in una diminuzione della quantità di ossigeno inspirato e quindi della quantità di ossigeno che raggiunge i diversi distretti cellulari (Beall 2007b), con conseguente alterazione delle funzioni metaboliche dell'organismo (Cogo 2009).

Nonostante ciò le popolazioni umane si sono insediate da decine di migliaia di anni in ambienti di alta quota, e i tre esempi più noti e numericamente rilevanti sono costituiti dalle popolazioni native degli altopiani etiopi, delle montagne del Tibet o delle Ande peruviane.

Ad oggi, circa 140 milioni di persone nel mondo vivono al di sopra dei 2500 m; di questi circa 17 milioni vivono oltre i 3500 m e 4 milioni al di sopra dei 4500 m (Penaloza and Arias-Stella 2007, Cogo 2009). Quindi per rispondere ad uno stress fisiologico per il quale non esistono adattamenti di tipo culturale, queste popolazioni hanno sviluppato caratteristiche ereditarie di riposta all'ipossia tali da poter permettere la vita in alta quota (Beall 2006).

1.2 Adattamenti fisiologici alla bassa pressione di ossigeno tipica dell'alta quota

1.2.1 Acclimatazione e adattamento

Quando un soggetto nativo di bassa quota si trasferisce ad alta quota presenta una sintomatologia caratterizzata da accorciamento del respiro con aumento della frequenza (tachipnea), stress respiratorio, fatica fisica e mentale, aumento del battito cardiaco, disturbi del sonno, emicrania, problemi digestivi ed eventuale perdita di peso.

Questi sintomi rappresentano una risposta fisiologica messa in atto dall'organismo per rispondere in modo immediato (effetti acuti) alla mancanza di ossigeno. Si definisce quindi *acclimatazione* una fase durante la quale l'organismo concentra la maggior parte dei suoi sforzi per raggiungere un equilibrio interno comparabile a quello mantenuto a livello del mare. Il prezzo da pagare per raggiungere tale equilibrio sono i sintomi sopra descritti che delineano un quadro clinico che va sotto il nome di "acute mountain sickness (AMS)". Il periodo necessario per il superamento di questa fase e la gravità dei sintomi variano da individuo a individuo e in alcuni casi possono insorgere patologie più gravi quali l'edema polmonare di alta quota (High Altitude Pulmonary Edema: HAPE) e l'edema cerebrale di alta quota (High Altitude Cerebral Edema: HACE).

I sintomi della AMS sono reversibili o con la discesa a quota inferiori o con il raggiungimento dello stato di equilibrio (Monge 1942, Grocott et al. 2007, Cogo 2009).

I principali meccanismi messi in atto dall'organismo per rispondere alla mancanza di ossigeno coinvolgono sia l'apparato respiratorio sia l'apparato cardio-circolatorio.

Nel prendere in considerazione l'apparato respiratorio, si osserva un aumento della ventilazione che consente un incremento dell'aria inspirata ed espirata nell'unità di tempo; questo fenomeno viene definito risposta ventilatoria ipossica (HVR, hypoxic ventilatory response) (Beall 2007b). Inoltre a livello polmonare si verifica una vasocostrizione dei capillari polmonari che determina una riduzione del calibro dei vasi ed un aumento della pressione polmonare (ipertensione polmonare) (Groves et al. 1993). Questo è anche causato da una diminuzione della sintesi di ossido nitrico (NO) in alta quota rispetto al valore medio al livello del mare (7,4 p.p.m. (parti per milione)) e ciò

favorisce ulteriormente la vasocostrizione polmonare (Duplain et al. 2000). L'ipertensione polmonare si verifica dopo pochi minuti di ipossia ma è reversibile al ritorno a bassa quota.

Questi tipo di cambiamenti permettono di portare più ossigeno ai tessuti aumentando la velocità con cui viene trasportato allo scopo di compensare l'ipossiemia creatasi (Cogo 2009).

Le alterazioni che si osservano a livello dell'apparato ematopoietico invece permettono di aumentare la quantità di ossigeno trasportato. Per fare ciò tuttavia è necessario più tempo, pertanto gli effetti sono rilevabili dopo qualche giorno.

La riduzione di ossigeno infatti stimola le cellule della corteccia renale a produrre eritropoietina (EPO), ormone glicoproteico che stimola le cellule del midollo osseo a produrre globuli rossi. L'incremento di EPO si ha dopo 6 ore di esposizione a 1800 m e i valori persistono alti per 24 ore. A quote più elevate, 2500-3000 m, i valori di EPO continuano a salire nell'arco delle prime 48-72 ore e dopo 5-10 giorni di permanenza in altitudine i valori di eritropoietina si stabilizzano su livelli basali. L'aumento di eritropoietina determina un innalzamento del valore di ematocrito, ossia del rapporto tra parte corpuscolata del sangue, costituita principalmente da globuli rossi, e la parte liquida, plasma, che a livello del mare ha valori compresi tra 42-52% per i maschi e 36-46% per le femmine. All'incremento nella percentuale di globuli rossi coincide un aumento della molecola in essi contenuta e deputata al trasporto di ossigeno: l'emoglobina (Hb). Ogni molecola di emoglobina può trasportare fino a quattro molecole di ossigeno e, grazie a un sistema di legami cooperativi, massimizza la sua affinità per l'ossigeno a livello alveolare mentre la minimizza a livello capillare per favorirne il rilascio. Il contenuto normale di Hb del sangue è tra 13,2 e 17,5 g/dL per gli uomini e tra 12,0 e 16,0 g/dL per le donne e (Cogo 2009).

Il rapido incremento della concentrazione di emoglobina che si osserva nei primi 2 giorni di arrivo in alta quota dipende da una riduzione del volume del plasma dovuto ad una maggiore perdita di liquidi (Sawka et al. 1996), mentre l'incremento che si osserva dopo alcune settimane di permanenza in alta quota dipende dall'aumentata sintesi di eritropoietina. Allo stabilizzarsi dei valori di eritropoietina coincide lo stabilizzarsi dei

valori di emoglobina. Per individui che restano ad una data altitudine per periodi inferiori ad 1 settimana il cambiamento nei livelli di globuli rossi è irrilevante e non contribuisce significativamente all'acclimatazione (Palmer 2010).

La saturazione di ossigeno (SpO_2) è una misura che varia anch'essa con l'altitudine.

Con il termine saturazione di ossigeno si intende la percentuale di molecole di emoglobina saturate (Ossiemoglobina: HbO_2) sul totale delle molecole di emoglobina presenti ($HbO_2 + Hb$).

$$SpO_2 = \frac{[HbO_2]}{[HbO_2 + Hb]} \times 100$$

A 3500 m la saturazione di ossigeno in un soggetto a riposo non ancora acclimatato è intorno al 90-92%, a 4500 m intorno all'80-84% (Cogo 2009).

Parametro	Variazione
Ematocrito (HTC)	↑
Emoglobina (Hb)	↑
Saturazione di Ossigeno (SaO_2)	↓
Ipertensione polmonare	✓
Sintesi di Ossido Nitrico (NO)	↓

Tab.1 In tabella sono riassunte tutte le modificazioni che un soggetto nativo di bassa quota subisce durante un soggiorno in alta quota.

Questi sistemi di risposta all'ipossia permettono a individui nativi di bassa quota che si recano in altitudine di acclimatarsi fino al raggiungimento di una fase di equilibrio e, come detto in precedenza, sono reversibili al rientro a quote inferiori.

Si può parlare pertanto di un *adattamento reversibile (o temporaneo)*.

Ma cosa succede in individui nativi e residenti in alta quota?

In questi individui, costantemente sottoposti a stress ipossico ed esposti per millenni all'azione della selezione naturale, alcuni sistemi di risposta alla bassa pressione di ossigeno si sono stabilizzati in modo permanente e vengono ereditati di generazione in generazione. In questi casi si può parlare di un *adattamento permanente* all'ipossia.

1.2.2 Popolazioni native di alta quota

I tre principali altopiani colonizzati dall'uomo sono l'altopiano tibetano, l'altopiano andino e l'altopiano etiopico. Per ragioni storiche legate specialmente alla facilità di compiere studi scientifici nelle zone Andine rispetto al Tibet o agli altopiani africani, le Ande sono state il primo sito in cui scienziati come i pionieri degli studi sull'adattamento umano in alta quota come Carlos Monge Cassinelli e Paul Beker hanno effettuato le prime ricerche (Monge 1942, Baker 1969). Alla fine degli anni '70 è diventato accessibile anche il Tibet e sono iniziati oltre che studi specifici su queste popolazioni, anche studi comparativi tra l'adattamento andino e quello tibetano che hanno permesso di concludere come le due popolazioni abbiano seguito percorsi evolutivi distinti mettendo in atto strategie di adattamento differenti e acquisendo caratteristiche fisiologiche distinte in grado di favorire il raggiungimento del medesimo scopo: rispondere ai bassi livelli di ossigeno (Beall 2006). Da questo punto di vista invece l'Etiopia è ancora scarsamente investigata.

In particolare, per quanto riguarda le popolazioni andine si nota principalmente un innalzamento delle variabili ematiche rispetto ai valori di riferimento delle popolazioni di bassa quota. La concentrazione di emoglobina, ad esempio, in individui residenti a bassa quota presenta valori compresi tra 12,0 e 16,0 g/dL per le donne e tra 13,2 e 17,5 g/dL per gli uomini, mentre nelle popolazioni andine i valori si sono stabilizzati su valori medi più elevati compresi tra i 17,8 g/dL e 19,2 g/dL rispettivamente in donne e uomini (Beall et al. 1998, Beall 2007b).

Questi valori medi sono superiori a quelli che si osservano in individui nativi di bassa quota che si sono acclimatati all'ambiente di alta quota dopo un soggiorno di media durata (16,3 mg/dL) (Grocott et al. 2007). Ad alti valori di Hb corrispondono alti valori di EPO che stimola la produzione dei globuli rossi. La concentrazione di EPO è significativamente maggiore nelle popolazioni andine (31,6 mIU/ml (Gonzales et al. 2009)) rispetto ai valori medi osservati a livello del mare (8,7 mIU/ml).

Questo sistema di risposta alla ridotta pressione di ossigeno deve essere finemente regolato per evitare l'insorgenza di patologie causate da un'eccessiva produzione di globuli rossi.

Carlos Monge nel 1942 descrisse per primo i sintomi di una patologia, definita appunto Morbo di Monge o Chronic Mountain sickness (CMS), caratterizzata da un'eccessiva eritrocitosi. Come detto precedentemente nella fase di acclimatazione si possono avvertire in sintomi di una forma acuta di mountain sickness (AMS) che tuttavia si allevieranno o con il rientro a quote inferiori o con il raggiungimento di un equilibrio interno dovuto all'adattamento. Se si raggiunge questo equilibrio anche i parametri ematici si stabilizzano.

Nel caso della CMS la permanente eccessiva eritrocitosi determina un incremento dell'ematocrito che può raggiungere valori medi del 60-70%, rispetto ad una soglia del 50% considerata non-patologica. A percentuali di ematocrito così elevate consegue un aumento della viscosità del sangue e il rischio di trombosi nei vasi. Questo si associa ad una serie di sintomatologie come grave ipossemia (ridotta quantità di ossigeno che circola nel sangue), moderata o grave ipertensione polmonare che può evolvere in insufficienza cardiaca ed altre patologie cardiaco-polmonari. In alcuni casi può provocare il decesso (Leon-Velarde and Mejia 2008).

Questa patologia non colpisce solamente individui nativi di bassa quota che non riescono ad adattarsi all'altitudine, infatti le prime popolazioni in cui fu studiata da Carlos Monge furono proprio le popolazioni peruviane native di alta quota (Monge 1942). Sulla base di queste osservazioni Monge definì questa malattia come una "mancanza di acclimatazione" o "perdita di acclimatazione".

È tuttavia interessante sottolineare che tale patologia è assente nelle popolazioni tibetane anch'esse native di alta quota. Questo dipende dal fatto che queste popolazioni hanno evoluto dei sistemi di risposta alla mancanza di ossigeno che coinvolgono principalmente l'apparato respiratorio.

Infatti, potenziando le variabili di tipo ventilatorio, questi individui sono in grado di aumentare i livelli di ossigeno in circolo (Beall 2007b) senza incrementare le percentuali di ematocrito rischiando di incorrere in patologie come la CMS.

I valori di ventilazione a riposo, ad esempio, sono circa raddoppiati rispetto a quelli riscontrati in individui di bassa quota (15,0 L/min e 7,5 L/min) e maggiori anche di quelli di individui nativi delle Ande peruviane (10,5L/min)(Beall 2007b).

Per capire quali di queste risposte alle basse pressioni di ossigeno messe in atto in modo differente da queste popolazioni possano essere considerate ereditabili di generazione in generazione, in alcuni lavori presenti in letteratura sono stati stimati i livelli di ereditabilità di questi fenotipi.

Per fare questo, non essendo chiare le basi genetiche associate ai fenotipi, sono stati utilizzati degli approcci di genetica quantitativa. Nei lavori di Beall (2000) e Beall et al. (2007b) sono riassunti i valori degli indici di ereditabilità (h^2) stimati per popolazioni Andine e Tibetane in relazione ad alcune caratteristiche fenotipiche. Nello specifico h^2 rappresenta la proporzione di varianza totale di un tratto fenotipico attribuibile ai legami di parentela (relazioni genetiche) tra gli individui in una popolazione.

Confrontando i valori di h^2 tra andini e tibetani relativi alla ventilazione a riposo e alla concentrazione di emoglobina è stata riscontrata una significativa ereditabilità dei livelli di emoglobina sia negli andini sia nei tibetani, sebbene nel primo caso si tratti di valori elevati, mentre nel secondo di valori contenuti. Relativamente alla ventilazione a riposo h^2 assume valori significativi solo nelle popolazioni tibetane (Beall 2000).

Pertanto sia le popolazioni andine sia le popolazioni tibetane hanno evoluto dei meccanismi di risposta all'ipossia ereditabili di generazione in generazione pur seguendo percorsi evolutivi differenti.

Per questo, negli ultimi anni l'attenzione si è spostata sullo studio delle basi genetiche che sottendono a queste caratteristiche, e soprattutto recentemente, con l'avvento delle nuove tecnologie di sequenziamento del genoma, questo argomento è stato oggetto di numerosi studi.

1.3 Basi genetiche dell'adattamento all'alta quota

Gli studi effettuati per identificare i geni coinvolti in questo tipo di adattamento si sono basati sia sullo studio dei geni candidati, a partire quindi da un'ipotesi biologica, sia sull'analisi dell'intero genoma senza ipotesi a priori su dove la variabilità associata ai fenotipi adattativi potesse essere collocata.

Lo studio dei geni candidati si basa sulla scelta a priori di geni che codificano per proteine note per essere potenzialmente coinvolte nel fenotipo oggetto di studio e la successiva tipizzazione in un numero abbastanza elevato di individui. Questo approccio, seppur limitato a regioni definite può essere considerato un metodo ad alta risoluzione poiché consente di analizzare la variabilità di intere sequenze identificando anche polimorfismi rari (Goldstein 2009). Prima dell'era genomica inferenze riguardanti la selezione naturale venivano effettuate utilizzando esclusivamente questo approccio. Ciò ha permesso di raggiungere importanti successi nella comprensione della storia evolutiva dell'uomo, come ad esempio l'associazione tra la variabilità al locus *LCT* e la persistenza della tolleranza al lattosio nell'individuo adulto (Bersaglieri et al. 2004) e la presenza di polimorfismi al locus *G6PD* associati ad una ridotta suscettibilità alla malaria (Tishkoff et al. 2001).

Nell'ultima decade questo approccio è stato applicato anche all'analisi dei geni candidati alla risposta all'ipossia (Moore et al. 2002, Rupert et al. 2003, Bigham et al. 2008, Stobdan et al. 2008). I risultati di queste ricerche non hanno permesso di definire delle associazioni chiare tra polimorfismi ai geni oggetto di studio e i fenotipi di risposta allo stress.

In particolare gli studi compiuti sulle popolazioni andine si possono riassumere in alcuni lavori su singoli geni e poche varianti geniche. Nonostante il suo ruolo centrale nella regolazione dell'omeostasi dell'ossigeno (Semenza 2007a) solo pochi studi hanno investigato la variabilità genetica al locus *HIF-1 α* in popolazioni di alta quota andine. In uno di questi studi il cDNA di *HIF-1 α* , dunque la sola regione codificante, di tre individui Quechua è stato sequenziato senza trovare varianti polimorfiche tra Quechua e nemmeno in confronto ad una sequenza di riferimento (Hochachka and Rupert 2003). Più interessante è il risultato di un recente studio intrapreso per misurare l'espressione

genica che possa predire l'insorgere di chronic mountain sickness, che ha mostrato come individui di alta quota mostrino una maggiore espressione dei geni regolati da HIF e minori livelli di quelli coinvolti nella glicolisi degli acidi tricarbossilici (Huicho et al. 2008). Passando ad altri geni candidati, un polimorfismo nel gene per l'angiotensin I converting enzyme (ACE, direttamente associato con la vasocostrizione e la regolazione della pressione sanguigna) è stato oggetto di studi di associazione in Andini di alta quota. Uno di questi studi mostra una mancanza di associazione tra genotipo e prestazioni cardiovascolari in individui di alta quota comparati con individui di bassa quota (Rupert et al. 2003). Un altro studio recente ha mostrato associazione tra uno specifico allele del locus ACE ed una maggiore saturazione di ossigeno arterioso in alta quota, ma in modo indipendente dallo stato di adattamento dell'individuo (nativi e non nativi di alta quota mostrano lo stesso grado di associazione) (Bigham et al. 2008). Studi su popolazioni andine aventi per oggetto il β 2-adrenergic receptor e il β -fibrinogeno basati sulla variabilità di alcuni SNP conosciuti a priori non hanno trovato segnali di associazione con l'adattamento alla quota (Rupert et al. 1999, Rupert et al. 2000).

I limiti di questi studi sono da ricercare non soltanto nella difficoltà di definire il corretto gene coinvolto in un fenotipo così complesso come quello dell'adattamento alla quota, ma anche nelle metodologie utilizzate per lo studio di questi geni, quasi esclusivamente indagati per singoli polimorfismi, e in un limitato numero di individui.

Recentemente l'accesso alle nuove tecnologie di sequenziamento ad elevato parallelismo (Mardis 2008, Shendure and Ji 2008) ha portato ad un incremento degli studi che coinvolgono l'intero genoma, in inglese "*genome-wide*".

In generale questi studi si basano sulla analisi di un numero elevato di polimorfismi di singolo nucleotide (SNP: *Single Nucleotide Polymorphisms*) scelti per essere rappresentativi di grosse regioni (tagSNPs) ed equamente distribuiti lungo tutto il genoma.

I principali approcci metodologici che utilizzano marcatori distribuiti lungo il genoma si basano sugli studi di associazione e sulla ricerca di loci che mostrano segni di selezione naturale.

Nel caso della ricerca delle basi genetiche dell'adattamento all'alta quota, gli studi di associazione adottano le strategie di analisi proprie degli studi caso-controllo condotti per scoprire le basi genetiche delle malattie complesse. Questo metodo prevede il confronto tra un campione "caso", cioè una popolazione di alta quota, e un campione "controllo" di bassa quota alla ricerca di differenze significative nelle frequenze alleliche di marcatori localizzati in regioni che, dunque, diventeranno candidate per uno studio più approfondito di sequenziamento.

Nel secondo caso i risultati della scansione dell'intero genoma vengono riassunti in distribuzioni empiriche basate su alcune statistiche che quantificano alcuni aspetti della variabilità genetica (*Summary statistics*). Queste statistiche sono in grado di rilevare deviazioni da condizioni di neutralità (test di neutralità) sulla base di differenze nelle frequenze alleliche (*Site frequency spectrum*), associazione non casuale degli alleli a due o più loci (*Linkage Disequilibrium, LD*) o struttura delle popolazioni (F_{ST}). Infatti, i loci presenti nelle regioni estreme delle distribuzioni (outliers) mostrano un pattern di variabilità genetica insolito rispetto al resto del genoma e ciò potrebbe dipendere dall'azione della selezione naturale.

Entrambi gli approcci permettono di rilevare varianti comuni responsabili del fenotipo oggetto di studio senza effettuare assunzioni a priori riguardo la natura delle varianti prese in considerazione. Tuttavia gli SNP scelti sono tra quelli con frequenza della variante più rara non inferiore al 5%, pertanto tutte le varianti a bassa frequenza saranno scartate a priori e non saranno rilevate nelle analisi successive. L'esclusione di questi polimorfismi rari può influenzare in modo rilevante lo spettro delle frequenze alleliche causando una interpretazione erronea dei risultati dovuta alla presenza di falsi positivi o falsi negativi (*ascertainment bias*) (Kreitman and Di Rienzo 2004, Akey 2009). Pertanto l'affidabilità dei risultati deve essere verificata sovrapponendo più studi e verificando la riproducibilità dei risultati (Akey 2009).

In conclusione quindi gli approcci basati sulla scansione dell'intero genoma attraverso tagSNPs mantengono un livello di risoluzione basso e costi elevati dovuti alla necessità di tecnologie ad elevato parallelismo.

Recentemente sono stati sviluppati dei metodi che permettono il sequenziamento di tutta la porzione codificante del genoma costituita dagli esoni, si parla infatti di *exome sequencing*, ossia il sequenziamento dell' "esoma"(Ng et al. 2010). Questo metodo è molto utile per lo studio di fenotipi determinati da polimorfismi che alterano la regione codificante del genoma, tuttavia esclude dall'analisi tutte le regioni regolatrici quali promotori, introni, UTR, a meno di piccole porzioni dovute ad una percentuale di errore nella fase di iniziale del processo, durante la quale vengono "catturate", dall'intero genoma, le regioni codificanti che verranno successivamente sequenziate.

Va ricordato che le regioni non codificanti del genoma svolgono in molti casi un ruolo fondamentale nella modulazione dell'espressione genica e nella manifestazione di fenotipi patologici o di adattamento all'ambiente (Tishkoff et al. 2007).

L'applicazione di questi metodi allo studio delle popolazioni native di alta quota ha portato recentemente a numerosi risultati che tuttavia vanno presi in considerazione tenendo conto delle problematiche sopra descritte.

La maggior parte di questi lavori su popolazioni native di alta quota si sono focalizzati esclusivamente sullo studio di popolazioni tibetane (Beall et al. 2010, Bigham et al. 2010, Simonson et al. 2010, Yi et al. 2010) mentre solo due studi successivi di Bigham et al. (2009 e 2010) hanno preso in considerazione anche popolazioni andine (Bigham et al. 2009, 2010).

L'approccio applicato da Bigham et al. (2009 e 2010) e da Simonson et al. (2010) appartiene alla tipologia di studi che cercano loci selezionati in ambiente di alta quota.

Tra i geni la cui variabilità risulta significativa vengono scelti quelli che mostrano maggiore attinenza con il fenotipo, ossia geni che appartengono a metabolismi regolati in ipossia.

Questi studi hanno portato all'identificazione del gene *EGLN1* come possibile gene associato a fenotipi di adattamento all'altitudine sia in popolazioni andine che tibetane. In particolare sia nello studio di Simonson et al. (2010) sia in quello di Bigham et al. (2010)

questo gene ha mostrato segnali significativi di associazione con la concentrazione di emoglobina sia in popolazioni andine che tibetane, seppur il pattern di variabilità a questo locus differisca nelle due popolazioni (Bigham et al. 2010), evidenziando ancora una volta i diversi percorsi evolutivi seguiti da Tibetani e Andini.

Restringendo l'analisi ai geni appartenenti a cicli metabolici direttamente coinvolti con la fisiologia in alta quota, i geni *PRKAA1* E *NOS2A* sono risultati i migliori candidati per l'adattamento negli andini, mentre *EPAS1* sembra essere il miglior candidato nei tibetani (Scheinfeldt et al. 2010).

A dimostrazione del fatto che non sempre i risultati di questi studi convergono nella stessa direzione, negli studi di Beall et al. (2010) e Yi et al. (2010), effettuati solo su popolazioni tibetane, il gene maggiormente associato all'adattamento all'alta quota risulta essere *EPAS1* (o *HIF-2 α*) e nello specifico le analisi di associazione con i fenotipi hanno mostrato associazioni significative del gene con bassi livelli di emoglobina, tipici delle popolazioni tibetane (Beall et al. 2010, Yi et al. 2010).

Il lavoro di Beall et al. (2010), che ha per oggetto solo il Tibet, appartiene invece alla categoria di studi che si basano su un approccio di ricerca dei marcatori che mostrano una significativa differenza di frequenza tra alta e bassa quota, ed al successivo sequenziamento ad alta risoluzione della regione identificata. L'analisi viene effettuata calcolando un χ^2 sulla base delle frequenze alleliche SNP per SNP, correggendo per eventuali effetti di stratificazione delle popolazioni.

In questo studio il gene *EPAS1*, che ha mostrato segnali significativi di associazione con l'alta quota dall'analisi GWAS, è stato successivamente sequenziato in campioni di alta quota tibetani. Questo ha permesso di identificare anche polimorfismi che non erano stati presi in considerazione dal GWAS e di effettuare per ciascuno i test di associazione con il fenotipo maggiormente correlato al ruolo svolto da questo gene, ossia la concentrazione di emoglobina.

I risultati di questo studio evidenziano la presenza di circa 26 polimorfismi intronici, in alto LD tra loro, il cui l'allele a frequenza maggiore risulta associato a basse concentrazioni di emoglobina in individui tibetani (Beall et al. 2010, Yi et al. 2010). Va ricordato che le

basse concentrazioni di Hb in queste popolazioni vengono interpretate come un fenomeno adattativo, nel senso che le popolazioni di queste regioni avrebbero raggiunto un equilibrio con la bassa pressione di ossigeno attraverso vie che non prevedono la necessità di aumentare il volume di globuli rossi che può portare ad effetti collaterali.

Un risultato simile è stato raggiunto anche nel lavoro di Yi et al. (2010) dove, utilizzando un approccio "*exome sequencing*", gli autori hanno identificato il gene *EPAS1* e hanno rilevato la presenza di un polimorfismo intronico, localizzato 4bp a monte dell'esone 6, che mostra anch'esso associazione significativa con la concentrazione di emoglobina (Yi et al. 2010). Va sottolineato che tale polimorfismo, essendo intronico, non sarebbe stato identificato a meno di quella piccola percentuale di errore nella fase di "cattura" sopra descritta.

Come è evidente, gli approcci descritti sopra sono metodi estremamente potenti nelle fasi iniziali di uno studio, tuttavia necessitano di essere seguiti da approcci di sequenziamento delle regioni identificate per poter ottenere risultati precisi che permettano di capire nel dettaglio quali polimorfismi siano effettivamente rilevanti per il fenotipo oggetto di studio (Akey 2009).

Accanto alle metodologie di analisi illustrate, esiste una sorta di via intermedia che può considerarsi un approccio gene candidato basato sulla conoscenza delle reti metaboliche. Questo permette di evidenziare un intero set di geni codificanti per proteine le cui interazioni definiscono il fenotipo di interesse. Grazie a metodi bioinformatici, che saranno descritti in seguito, è possibile delineare un quadro di geni candidati da studiare, basandosi sulle informazioni provenienti da più database pubblici. Questi database, interconnessi tra loro, permettono di unire le informazioni di più studi effettuati in ambiti differenti (profili di espressione, reti metaboliche, interazioni proteiche, studi funzionali, etc.) e di scegliere i geni candidati tra quelli che mostrano una maggiore quantità di evidenze. Come descritto successivamente, questi metodi ci hanno suggerito la via metabolica di HIF come candidata per lo studio delle basi genetiche dell'adattamento all'alta quota.

1.4 La via metabolica di HIF

L'esposizione dell'organismo umano a condizioni di ipossia causa molti cambiamenti sia a livello sistemico che cellulare al fine di minimizzare gli eventuali danni e ripristinare l'ossigenazione adeguata in ogni distretto. Queste modificazioni sono a carico di una complessa macchina regolatrice, sensibile alla concentrazione di ossigeno, composta da molte componenti molecolari che si attivano in tempi e modi diversi (Smith et al. 2008).

I principali regolatori dell'omeostasi dell'ossigeno, del suo trasporto e dell'adattamento a condizioni di ipossia sono i fattori di trascrizione HIF (Hypoxia Inducible Factors).

Queste molecole sono eterodimeri composti da una subunità α il cui destino è ossigeno dipendente e una subunità β , espressa in modo costitutivo (Semenza 2007a, Kaelin and Ratcliffe 2008, Majmundar et al. 2010).

I metazoi possiedono tre isoforme di HIF- α di cui HIF-1 α e HIF-2 α (anche noto come EPAS1) sono quelli più simili da un punto di vista strutturale e meglio caratterizzati.

HIF-3 α (o IPAS) esiste in più forme, in base allo splicing alternativo subito, alcune delle quali inibiscono l'attività di HIF-1 α e HIF-2 α (Kaelin and Ratcliffe 2008, Majmundar et al. 2010). HIF-1 α è espresso in modo ubiquitario in tutte le cellule, mentre HIF-2 α e HIF-3 α sono selettivamente espressi solo in certi tessuti quali ad esempio le cellule dell'endotelio vascolare, gli pneumociti di tipo II, le cellule renali interstiziali, gli epatociti e le cellule mieloidi (Majmundar et al. 2010).

In presenza di ferro e concentrazione normale di ossigeno, la subunità α di HIF viene continuamente sintetizzata e rapidamente degradata mediante ubiquitinazione (Semenza 2007b, Kaelin and Ratcliffe 2008). La degradazione inizia con l'idrossilazione di due residui di prolina altamente conservati localizzati nei domini ODD (oxygen dependent degradation domain) e N-TAD (N-terminal Transactivation domain) di HIF- α (Fig.1).

L'idrossilazione avviene a carico di enzimi Prolil Idrossilasi (PHD, dette anche EGLN) la cui attività è ossigeno dipendente. Esistono tre forme di questi enzimi derivanti da geni paraloghi, ossia originatesi per duplicazione genica: PHD1, PHD2 e PHD3 codificate dai geni *EGLN1*, *EGLN2* e *EGLN3*. Le prolil idrossilasi maggiormente espresse in ipossia sono PHD2, a livello citoplasmatico, e PHD3 localizzata sia a livello citoplasmatico sia a livello nucleare (Kaelin and Ratcliffe 2008).

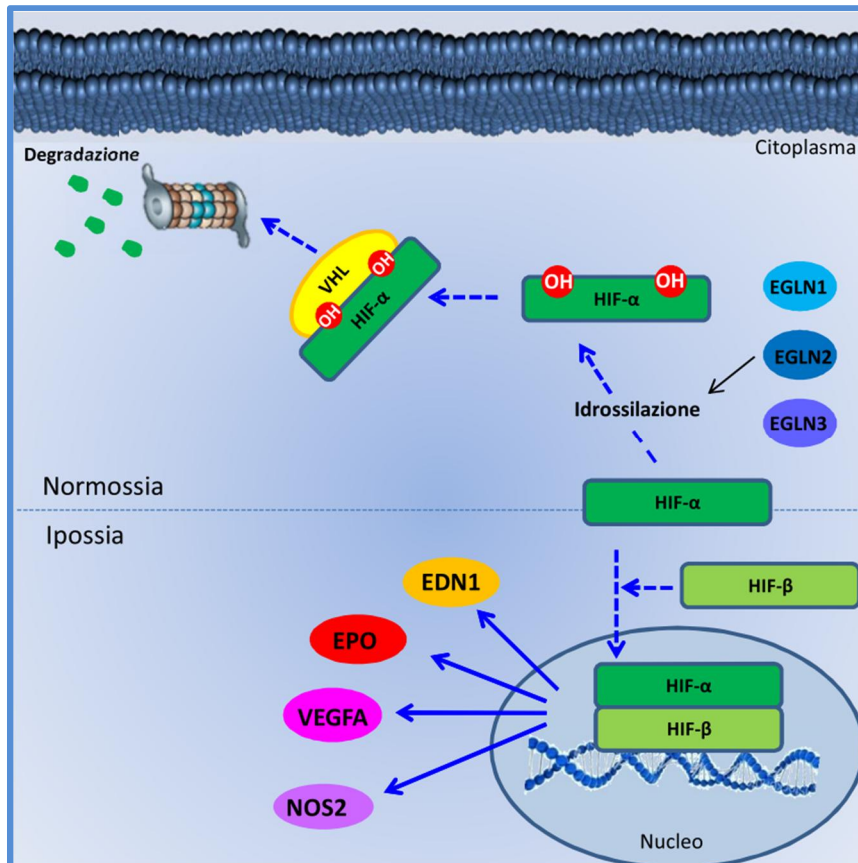


Fig.1 Il ciclo metabolico di HIF

L'idrossilazione di entrambi i residui di prolina determina la formazione di un sito di legame per la proteina VHL (Von Hippel-Lindau tumour suppressor protein). VHL servirà poi da substrato per un altro enzima, una E3 ubiquitina-ligasi (E3 ubiquitin protein ligase) che porterà a degradazione rapida il complesso attraverso un sistema ubiquitina-proteosoma.

Contemporaneamente all'idrossilazione delle due proline, viene anche β -idrossilato un residuo di Asparagina localizzato nel dominio C-TAD (C-terminal Transactivation Domain) di HIF-1 α e HIF-2 α . Questa volta l'idrossilazione avviene da parte dell'enzima FIH-1 (factor inhibiting HIF-1) sempre ossigeno dipendente. Questa modificazione impedisce il reclutamento dei co-attivatori trascrizionali CBP/p300, inibendo quindi l'attività di transattivatore di HIF (Kaelin and Ratcliffe 2008). FIH-1, rispetto a PHD, resta attivo anche in presenza di bassi livelli di ossigeno quindi è in grado di controllare la degradazione di HIF anche in condizioni di moderata-lieve ipossia (Kaelin and Ratcliffe 2008).

Con queste due differenti, ma complementari strade, la subunità α di HIF viene degradata e non potrà interagire con la subunità β per formare il fattore di trascrizione attivo HIF.

Quando la concentrazione di ossigeno scende sotto i livelli normali, viene inibita l'attività dei suddetti enzimi, PHDs e FIH-1 e la degradazione proteosomale di HIF- α è rallentata. HIF- α viene così rapidamente accumulato dimerizza con la subunità β costitutivamente espressa e trasloca a livello nucleare dove, dopo il legame con il complesso co-attivatore p300/CBP, attiva la trascrizione di numerosi geni (Semenza 2001, Semenza 2007b, Majmundar et al. 2010) i cui prodotti proteici sono coinvolti in diversi processi tra cui l'angiogenesi (*VEGFA*), l'eritropoiesi (*EPO*) e la glicolisi (*GLUT*) (Semenza 2004, Kaelin and Ratcliffe 2008).

La conservazione lungo la filogenesi dei metazoi di tutte e tre le isoforme sottolinea l'importanza di queste tre proteine che svolgono un ruolo essenziale nella mediazione della risposta allo stress ipossico (Kaelin and Ratcliffe 2008, Lendahl et al. 2009).

In particolare HIF-1 α e HIF-2 α sono coinvolti in numerosi processi metabolici.

Studi di regolazione selettiva della trascrizione hanno portato all'identificazione di geni target esclusivi di HIF-1 α , altri controllati solo da HIF-2 α , mentre alcuni geni risultano regolati dall'attività di entrambe le proteine (Hu et al. 2003, Raval et al. 2005, Holmquist-Mengelbier et al. 2006, Kaelin and Ratcliffe 2008). Ad esempio HIF-1 α controlla il metabolismo del glucosio, mentre il metabolismo degli acidi grassi è regolato principalmente da HIF-2 α . Nel caso invece dei processi di angiogenesi l'attività dei fattori di crescita vascolare è controllata da entrambe le molecole (Majmundar et al. 2010). Tutto ciò dipende soprattutto dal tipo di tessuto in cui le proteine si esprimono. E' stato dimostrato che in cellule in cui oltre ad HIF-1 α viene espresso anche HIF-2 α , questo presenta attività ridotta, rispetto all'isoforma HIF-1 α (Hu et al. 2006, Majmundar et al. 2010).

Nei capitoli successivi di questa tesi verranno descritti i metodi di selezione dei geni candidati che saranno studiati in questo progetto. La maggior parte di questi appartengono al ciclo metabolico appena descritto. In particolare in questa tesi saranno descritti i risultati del sequenziamento dei geni *EPO*, *VHL* ed *EPAS1* dai quali siamo partiti per lo studio delle basi genetiche dell'adattamento alle basse pressioni di ossigeno.

EPO è un gene target di HIF, pertanto si trova a valle del ciclo metabolico ed è coinvolto nella produzione di eritropoietina che, come abbiamo visto nei paragrafi 1.2.1 e 1.2.2, svolge un ruolo importante nella regolazione della quantità di ossigeno in circolo, e la sua sintesi viene appunto stimolata in condizioni di bassa pressione di ossigeno.

La proteina VHL invece si trova a monte del ciclo metabolico di HIF e svolge un ruolo chiave nella regolazione della degradazione di questo fattore di trascrizione. Sono state riscontrate in letteratura numerose evidenze che dimostrano che alterazioni della struttura proteica di VHL causate da mutazioni a livello di alcune regioni del gene sono presenti in numerose patologie tumorali dove si assiste ad una sovra-regolazione di questo ciclo metabolico.

Per quanto concerne *EPAS1* è stato discusso il suo ruolo nei paragrafi 1.3 e 1.4 soprattutto in relazione a popolazioni tibetane e al suo coinvolgimento nella regolazione della concentrazione di emoglobina. Riguardo questo gene saranno presentati i risultati di un'analisi preliminare di sequenziamento di alcune regioni.

1.5 EPO

L'eritropoietina è un ormone glicoproteico in grado di promuovere la differenziazione degli eritrociti (globuli rossi) e di regolarne la concentrazione nel sangue. La scoperta di questa proteina e del suo ruolo fondamentale per l'apporto di ossigeno ai tessuti risale proprio a studi effettuati in ambiente di alta quota intorno alla metà del XIX secolo .

Paul Bert, professore di fisiologia all'università Sorbonne di Parigi e il suo collaboratore Denis Jourdanet studiando animali di alta quota (4000 m) rilevarono in questi una maggiore concentrazione di globuli rossi associata ad una migliore disponibilità di ossigeno ai tessuti rispetto ad animali di bassa quota. Supposero inoltre che questa caratteristica, a cui diedero il nome di "poliglobulia", fosse ereditaria.

Questa supposizione fu successivamente contestata dagli studi successivi effettuati dall'anatomista francese Francois-Gilbert Viault che, a seguito di un periodo trascorso sulle montagne del Perù, rilevò egli stesso un incremento dei globuli rossi nel suo sangue e in quello dei suoi accompagnatori (Dr. Mayorga, alcuni minatori, un cane, un gallo e un lama). Viault concluse che il processo di eritropoiesi (produzione di globuli rossi) viene stimolato in modo acuto quando la quantità di ossigeno nel sangue è ridotta. Questi dati furono confermati ulteriormente da studi successivi che portarono a definire l'eritropoiesi un processo con una regolazione ormonale e nel 1948 fu coniato il termine *ERITROPOIETINA* per definire il fattore ematopoietico che regola tale processo (Jelkmann 2007).

L'eritropoietina (*UniProt ID: P01588*) è una proteina costituita da 193 amminoacidi e codificata dal gene *EPO* la cui espressione tessuto specifica è dipendente dalle fasi dello sviluppo. Infatti nella fase fetale la proteina viene prodotta principalmente dagli epatociti, nel fegato, mentre nell'adulto il principale sito di espressione sono i reni (Ebert and Bunn 1999). Il momento in cui avviene questo passaggio è specie-specifico e nonostante ciò il fegato mantiene la capacità di sintetizzare la proteina. In particolare, in condizioni normali, il rapporto tra i livelli di espressione della proteina nei reni e nel fegato è di 9:1, mentre in condizioni di ipossia grave è stato stimato che la produzione di EPO a livello epatico ammonta a circa il 33% della quantità totale (Rankin et al. 2007).

Il principale stimolo che regola la sintesi di EPO è appunto l'ipossia tissutale e i primi studi effettuati per isolare il fattore responsabile del controllo della sintesi proteica hanno portato all'identificazione del fattore di trascrizione HIF-1 α in linee cellulari Hep3B del tessuto epatico umano (Semenza and Wang 1992). Nonostante ciò studi recenti hanno dimostrato che la sintesi di EPO è regolata principalmente da HIF-2 α sia a livello epatico sia a livello delle cellule renali (Rankin et al. 2007, Castrop and Kurtz 2010, Paliege et al. 2010).

Questo è probabilmente dovuto ad una maggiore affinità di HIF-2 α per un elemento di controllo della trascrizione (enhancer) definito Hypoxia Responsive Element (HRE) localizzato nella porzione 3' del gene *EPO* (Semenza et al. 1991, Rankin et al. 2007).

Il gene *EPO* (2901bp) (*EntrezGene ID: 2056*) è localizzato sul braccio corto del cromosoma 7 (7p22) ed è composto di 5 esoni. Non sono noti polimorfismi associati a patologie nella regione codificante del gene, e tale regione, come dimostrato anche nel nostro studio di evoluzione molecolare (vedi capitolo 7) , è generalmente molto conservato lungo la filogenesi (Wen et al. 1993). La regolazione della trascrizione del gene non dipende esclusivamente dal promotore, ma anche dalla presenza di un sito di legame (sequenza consenso TACGTGCT) per il complesso HIF- α -HIF- β in una sequenza di 259bp (HRE) localizzata immediatamente a valle della regione 3'UTR del gene (Semenza and Wang 1992, Percy et al. 1997).

1.6 VHL

VHL è l'acronimo di *Von Hippel Lindau*, una malattia neoplasica (OMIM ID: 193300) che determina lo sviluppo di una serie di tumori maligni e benigni in diversi organi e sistemi e prende il nome dai suoi scopritori: l'oftalmologo tedesco Eugen Von Hippel e il patologo svedese Arvid Lindau. Lo studio di questa malattia con ereditarietà autosomica dominante ha portato all'identificazione e alla caratterizzazione del gene *VHL*, le cui mutazioni causano la malattia stessa.

La comprensione del ruolo biologico svolto da VHL nella risposta cellulare all'ipossia rappresenta un importante esempio di come la ricerca nel campo delle scienze di base abbia portato nozioni fondamentali per la cura delle malattie nell'uomo.

La scoperta del gene *VHL* e conseguentemente della proteina da esso codificata risalgono a studi effettuati intorno alla fine degli anni '80. Studi recenti avevano dimostrato la cosiddetta *teoria dei due colpi o teoria di Knudson (Knudson 1971)*, secondo cui affinché si manifesti una patologia tumorale è necessario che entrambe le copie di un gene oncosoppressore siano inattivate a causa di due eventi mutazionali indipendenti. Secondo questa teoria l'insorgenza di patologie tumorali, non ereditarie, presenta carattere sporadico poiché affinché si manifesti la malattia è necessario che avvengano due mutazioni (eventi rari) a carico dello stesso gene e nella stessa cellula. Nelle forme ereditarie di tali patologie una delle due copie del gene viene ereditata con una mutazione, pertanto sarà sufficiente che si manifesti la seconda mutazione a carico dello stesso gene per inattivare la funzione della proteina codificata e determinare lo sviluppo di patologie tumorali.

Alla luce di questa teoria e dal momento che una delle prime manifestazioni della sindrome VHL era in comune con individui che presentavano forme sporadiche della stessa malattia si è ipotizzato che il gene responsabile di entrambe le patologie fosse appunto un gene oncosoppressore. Studi di citogenetica e analisi RFLP (Restriction Fragment Length Polymorphisms) in famiglie affette da carcinoma delle cellule renali (RCC) portarono all'identificazione di anomalie a livello del braccio corto del cromosoma 3 (Clark and Cookson 2008). La scoperta del gene *VHL* risale al 1993 al termine di una serie di studi di clonaggio posizionale svolti da diversi gruppi di ricerca (Latif et al. 1993).

Il gene *VHL* (*EntrezGene ID: 7428*), è lungo 10444 bp ed è localizzato nella regione 3p25.3. È costituito da 3 esoni che codificano per una proteina di 213 amminoacidi (variante 1, p30) (*UniProt ID: P40337*). Questa proteina è localizzata principalmente nel citoplasma di cellule renali, ma è stata rilevata anche a livello nucleare e associata alle membrane cellulari. Esiste anche un'isoforma proteica di 160 amminoacidi (variante 2, p19) la cui trascrizione dipende dalla presenza di un sito di inizio trascrizione più interno (Met54). Questa proteina è espressa equamente nei distretti citoplasmatici e nucleari mentre non è presente sulle membrane cellulari (Iliopoulos et al. 1998). Entrambe le proteine presentano attività di oncosoppressori, sono presenti in molti organismi pluricellulari e la loro sequenza è generalmente conservata (Nordstrom-O'Brien et al. 2010).

I primi studi effettuati per chiarire il ruolo di questa proteina evidenziarono che in linee cellulari mancanti del gene *VHL*, si aveva una alterata regolazione dell'espressione di due proteine la cui trascrizione è attivata da condizioni di ipossia cellulare (Vascular Endothelial Growth Factor (VEGF) e Glucose Transporter 1 (GLUT1)). Reintroducendo in tali cellule la forma wild-type del gene la regolazione della sintesi di queste proteine ritornava a livelli normali. Questo suggerì che VHL fosse in qualche modo importante nei processi metabolici di risposta all'ipossia (Clark and Cookson 2008).

Studi successivi dimostrarono che VHL svolge un ruolo fondamentale all'interno della cascata metabolica di HIF- α , fungendo da principale regolatore della degradazione di quest'ultimo (Clark and Cookson 2008).

Come descritto nel paragrafo 1.4, in condizioni normali di ossigeno HIF- α subisce l'idrossilazione di due residui di prolina da parte dell'enzima PHD (prolil-idrossilasi). Questa idrossilazione determina la formazione di due siti di legame per VHL il quale attraverso l'interazione con il complesso E3-ubiquitina-ligasi porterà HIF- α alla degradazione attraverso un sistema ubiquitina-proteosoma (Fig.2).

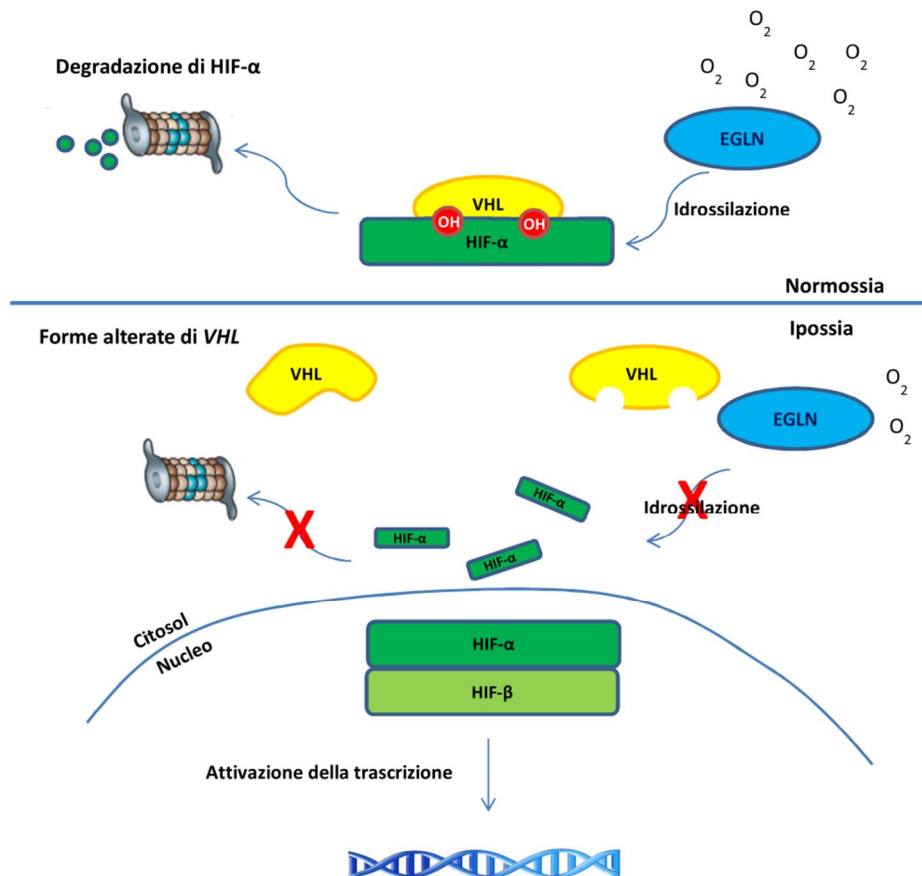


Fig.2 Ruolo della proteina Von Hippel Lindau (VHL) nella via metabolica di HIF. La prima parte della figura si riferisce alla degradazione della subunità HIF- α in condizioni di normossia via idrossilazione ad opera delle Prolil Idrossilasi (PHD) codificate dai geni *EGLN* (vedere tabella 7 capitolo 6, geni candidati). La seconda parte della figura mostra come la mancata degradazione della subunità HIF- α porti alla dimerizzazione con HIF- β nel nucleo in condizioni di ipossia.

In presenza di forme alterate, non funzionali della proteina VHL, viene in un certo senso “mimato” un ambiente di ipossia per cui non si assiste all’interazione di VHL con HIF- α e quest’ultimo accumulandosi entra a livello del nucleo e interagendo con altri fattori induce la trascrizione di geni che svolgono un ruolo importante nella risposta cellulare all’ipossia. Questo determina una attività eccessiva di HIF- α in condizioni normali di ossigeno in cui non sarebbe richiesta provocando alterazioni di tutti i sistemi controllati da molecole la cui trascrizione è appunto dipendente da questo fattore (Clark and Cookson 2008, Kaelin 2008, Nordstrom-O'Brien et al. 2010) (Fig.2).

Sono noti anche altri processi nei quali VHL svolge un ruolo importante quali l'orientamento e stabilità dei microtubuli, la regolazione della senescenza e la trasmissione dei segnali citochinici (Nordstrom-O'Brien et al. 2010).

La maggior parte delle mutazioni note nella sequenza codificante del gene sono associate all'insorgenza di patologie tumorali poiché causano una mancata o alterata interazione tra VHL e HIF- α (Kim and Kaelin 2004, Nordstrom-O'Brien et al. 2010). Queste mutazioni comprendono sia sostituzioni amminoacidiche sia vere e proprie aberrazioni cromosomiche. In particolare la regione che codifica per i residui minimi di interazione con HIF-1 α (aminoacidi da 91 a 113), localizzata nell'esone 1 del gene, sembra essere un hotspot mutazionale associato all'insorgenza di patologie tumorali (Tanimoto et al. 2000, Nordstrom-O'Brien et al. 2010). Anche l'esone 3 del gene svolge un ruolo fondamentale per il corretto funzionamento della proteina, poiché codifica per il dominio di VHL che interagisce con le proteine Elongin B e C del complesso E3-Ubiquitina-ligasi che porta HIF- α alla degradazione (*UniProt*: <http://www.uniprot.org/uniprot/P40337>). Sono infatti note mutazioni a livello di questa regione del gene che determinando difetti nell'interazione dei complessi proteici causando l'insorgenza di neoplasie (Nordstrom-O'Brien et al. 2010). Per quanto riguarda le regioni non codificanti del gene, studi effettuati in individui affetti da carcinoma delle cellule renali hanno evidenziato associazioni tra l'insorgenza del tumore e una maggiore metilazione della regione 5' di *VHL* (Herman et al. 1994, Banks et al. 2006, McRonald et al. 2009, Young et al. 2009). Questo evidenzia la possibilità di controlli epigenetici in grado di inattivare la corretta espressione della proteina con conseguenze a carico dell'intero ciclo metabolico di HIF- α .

Alterazioni nel funzionamento di VHL sono state riscontrate anche in individui policitemici. In particolare la sostituzione amminoacidica Arg200Trp dovuta ad una sostituzione nucleotidica nell' esone 3 del gene (dbSNP ID: *rs28940298*), è stata associata ad una malattia autosomica recessiva caratterizzata da eccessiva eritrocitosi e diffusa principalmente nella popolazione Chuvash della Russia centrale (Semenza 2009), da cui appunto il nome "*policitemia di Chuvash*". Anche in questo caso la mutazione riduce l'interazione di VHL con la proteina HIF-1 α idrossilata determinando una riduzione della

degradazione di quest'ultimo ed una elevata trascrizione del gene *EPO* anche in condizioni di normossia (Semenza 2009). I pazienti affetti da policitemia di Chuvash mostrano inoltre livelli di ventilazione a riposo elevati, vasocostrizione, ipertensione polmonare e incrementi del battito cardiaco, tutti sintomi comuni alle fasi di acclimatazione all'alta quota (Smith et al. 2006, Droma et al. 2008, Semenza 2009).

1.7 EPAS1 (HIF-2 α)

Il gene *EPAS1* (*EntrezGeneID: 2034*), acronimo di *endothelial pas domain 1*, codifica per l'isoforma -2 α di HIF, infatti è comunemente noto anche con il nome di *HIF-2a*.

Il gene è lungo 89302bp ed è localizzato sul braccio corto del cromosoma 2. È costituito da 16 esoni codificanti per una proteina di 870 amminoacidi (*UniProt ID: Q99814*). Il suo importante ruolo nel controllo della risposta adattativa all'ipossia nelle popolazioni tibetane è stato messo in luce da più di uno studio genomico, come descritto nel paragrafo 1.3.

2. SCOPO DEL PROGETTO

Questo studio ha come scopo principale definire il modo in cui la variabilità genetica ad alcuni loci candidati influisca sull'adattamento alla bassa pressione di ossigeno in popolazioni native di ambienti di alta quota delle Ande peruviane.

Dal punto di vista dei geni da studiare abbiamo ritenuto di definire una strategia per individuare i loci più indicati basata sulla letteratura e sull'ausilio di mezzi bioinformatici sviluppati appositamente per questi scopi. Lo studio dell'evoluzione molecolare di questi geni attraverso un approccio interspecifico è stato intrapreso per fornire importanti informazioni circa le regioni geniche più o meno conservate e, dunque, interessanti per diversi aspetti dello studio sulle popolazioni umane.

Il raggiungimento dello scopo principale di questo lavoro ha richiesto la raccolta di materiale biologico di popolazioni native da cui estrarre il DNA per effettuare gli studi genetici, insieme alle misurazioni di importanti caratteristiche fenotipiche legate all'adattamento alle basse pressioni di ossigeno, come variabili ematiche e polmonari. Allo studio di associazione genotipo/fenotipo compiuto all'interno del campione di alta quota, è stato affiancato un confronto della variabilità genetica tra alta e bassa quota. Quest'ultima tipologia di analisi è stata possibile grazie al sequenziamento degli stessi geni in un campione di individui di bassa quota che condividono con i nostri una simile storia demografica.

3. RACCOLTA E DESCRIZIONE DEL CAMPIONE POPOLAZIONISTICO

3.1 I campioni di alta quota

Come detto in precedenza, nonostante le popolazioni andine siano le prime popolazioni studiate in relazione all'adattamento all'alta quota, oggi gli studi si sono spostati sulle popolazioni dell'altopiano tibetano lasciando in sospeso molte domande relative alle caratteristiche di adattamento alle basse pressioni di ossigeno evolute appunto dalle popolazioni dell'altopiano andino.

Da alcuni anni il nostro gruppo di ricerca si occupa dello studio della variabilità genetica di popolazioni andine in relazione ad aspetti quali lo studio della demografia di queste popolazioni (Fuselli et al. 2003) e lo studio di loci di interesse farmacogenetico (Fuselli et al. 2007).

Per questi motivi abbiamo deciso di allargare lo spettro di conoscenze legato alla variabilità genetica di questa area geografica considerando in questo studio popolazioni composte da individui nativi e residenti nelle regioni di alta quota delle Ande peruviane.

Ai fini dello svolgimento di questo lavoro è stato necessario raccogliere materiale biologico, nello specifico si è trattato della raccolta di saliva, da cui estrarre il DNA ed effettuare una serie di misurazioni di parametri fisiologici per le successive analisi di associazione genotipo-fenotipo.

3.1.1 Siti di campionamento

Il campionamento è stato effettuato tra marzo e maggio del 2009 in tre siti distinti delle Ande peruviane scelti per la loro differente altitudine: Cusco 3326 m s.l.m., Cusibamba 3640 m s.l.m. e Puno 3830 m s.l.m. (Fig.3).



Fig.3 I siti di campionamento

A Cusco abbiamo lavorato presso l'Univesidad Nacional de San Antonio Abad del Cusco (UNSAAC) grazie alla collaborazione della professoressa Rosa Pacheco, della dottoressa Elsa Aguilar e di alcuni studenti.

Lavorare presso i locali della struttura universitaria ha permesso di velocizzare tutte le operazioni di analisi dei campioni biologici, di ridurre i rischi di contaminazione e di malfunzionamento delle attrezzature.

Il buon livello di istruzione dei soggetti che si sono prestati al campionamento ha reso più facile la descrizione del progetto e delle metodiche utilizzate, permettendo così diminuire la durata del campionamento ed effettuare, con un buon livello di precisione, le misurazioni fenotipiche (Fig.4).



Fig.4 Raccolta di materiale biologico e misurazioni fenotipiche a Cusco (3326 m s.l.m.): Universidad Nacional de San Antonio Abad del Cusco (UNSAAC).

Il secondo sito di campionamento è Cusibamba (3640 m s.l.m.), un piccolo villaggio localizzato a qualche ora da Cusco dove la gente vive senza luce elettrica ed acqua corrente. L'economia si basa esclusivamente sulla pastorizia e sull'agricoltura.

La raccolta dei campioni è stata effettuata presso i locali del salone comunale del villaggio. Non potendo lavorare in condizioni ottimali di igiene ambientale abbiamo utilizzato tutte le misure di precauzione possibili per minimizzare l'eventuale contaminazione dei campioni e della strumentazione preservando sempre la salute degli individui. Inoltre abbiamo tarato e corretto alcune misurazioni effettuate in quanto la strumentazione ha risentito di alcune variabili ambientali estreme come, ad esempio, la temperatura.

La collaborazione con i soggetti che si sono prestati al campionamento è stata complicata da motivi linguistici e culturali. Infatti gli abitanti di Cusibamba non parlano spagnolo ma Quechua, antica lingua locale che deriva dal Proto-Quechua che è stata la prima lingua che si è diffusa nella regione delle Ande. Per questo motivo con noi era presente un interprete, tuttavia anche lo scarso livello culturale ha reso più difficile la spiegazione e la comprensione delle metodiche utilizzate (Fig.5).



Fig.5 Raccolta di materiale biologico e misurazioni fenotipiche a Cusibamba (3640 m s.l.m.): salone comunale del villaggio.

Terzo ed ultimo sito di campionamento è Puno 3830 m s.l.m., sulle rive del lago Titicaca che funge da controverso confine tra Perù e Bolivia nella zone sud del paese. A Puno abbiamo lavorato al Banco de Sangre dell'Hospital Regional grazie alla collaborazione della professoressa Roxana Medina dell'Università di Puno e della dottoressa Vicky Cristina Gonzalez responsabile dell'ambulatorio di raccolta del sangue dell'ospedale. Trattandosi di un ospedale regionale è stato possibile raccogliere materiale ed informazioni di individui di Puno, ma anche di villaggi vicini di altitudini superiori ai 3830 m s.l.m.. Inoltre la posizione strategica della città, al confine tra Perù e Bolivia, ha permesso di campionare individui di gruppo etnico sia Quechua (tipico delle Ande peruviane) che Aymarà (maggiormente presente nella zona andina boliviana).

Lavorare presso i locali del Banco de Sangre ha semplificato le misurazioni di variabili ematiche e ha permesso di effettuare anche l'analisi della glicemia che non è stato possibile rilevare nei soggetti provenienti dagli altri siti di campionamento sopra descritti (Fig.6).



Fig.6 Raccolta di materiale biologico e misurazioni fenotipiche a Puno (3830 m s.l.m.): Banco del sangue presso l' Hospital Regional.

3.1.2 Procedura di campionamento, misurazioni fenotipiche e strumentazione

Il progetto è stato approvato dal comitato etico dell'Università di Ferrara e tutti gli individui maggiorenni che si sono prestati al campionamento sono stati informati circa lo scopo e le fasi dello studio, e la segretezza dei dati, tramite lettura, spiegazione e sottoscrizione di un modulo di consenso informato scritto. Inoltre ad ogni volontario è stato assegnato un codice identificativo come previsto dalle procedure che tutelano la privacy.

Per tutti gli individui sono state raccolte informazioni relative alla genealogia del padre e della madre fino a tre generazioni precedenti registrando i relativi luoghi di nascita. Questo ha permesso di avere informazioni importanti per le successive analisi dei dati, relative ai gradi di parentela degli individui campionati e all'eventuale presenza di

individui non nativi nel cui genoma coesistono componenti provenienti da pool genici di popolazioni differenti, nel caso specifico probabilmente di origine nativa americana ed europea.

Successivamente alla raccolta delle informazioni e dei consensi sono state effettuate le misurazioni fenotipiche: altezza, peso, perimetro del torace, pressione sanguigna, ritmo cardiaco, temperatura corporea, ossimetria (misurazione della percentuale di molecole di emoglobina saturate di ossigeno mediante una procedura non invasiva che sfrutta la diversa lunghezza d'onda di assorbimento della luce di molecole saturate e non saturate), prelievo sanguigno.

Il perimetro del torace è stato misurato con tre modalità differenti: a riposo chiedendo al soggetto di respirare normalmente, al massimo di una inspirazione chiedendo al volontario di trattenere l'aria per qualche secondo ed infine al massimo di una espirazione quando l'individuo aveva completamente svuotato i polmoni dall'aria.

La pressione sanguigna, sia sistolica che diastolica (in mmHg), è stata misurata con uno sfigmomanometro digitale (OMRON M2 Compact Upper arm Blood Pressure Monitor).

Lo stesso strumento misura il ritmo cardiaco (in BPM).

La temperatura corporea è stata rilevata con un termometro digitale a infrarossi (PRESTOTERM CORMAN) appoggiato per 5 secondi alla tempia del soggetto.

L'ossimetria (% SpO₂) è stata determinata grazie ad un ossimetro da dito (Drive Medical Fingertip Pulse Oximeter "pulsOx").

Le misurazione sopra descritte di pressione sanguigna, battito cardiaco, temperatura corporea ed ossimetria sono state effettuate almeno tre volte e i valori finali registrati sono stati ottenuti facendo la media delle tre rilevazioni.

Dopo aver misurato tutte queste variabili corporee, è stato prelevato del sangue venoso al soggetto per il calcolo dell'ematocrito, dell'emoglobina, del gruppo sanguigno e, solo per il sito di Puno, della glicemia.

La metodica seguita per ottenere il valore di ematocrito (rapporto tra il volume occupato dagli eritrociti rispetto al volume del sangue intero dopo centrifugazione, si esprime in %) è la seguente: si riempiono due capillari di vetro e si centrifugano a 10-15 mila g per 5 minuti. Ciò provoca la separazione dei diversi componenti del sangue intero e, tramite la sovrapposizione del capillare su delle tabelle graduate, si ricava il valore di ematocrito corrispondente.

La concentrazione di emoglobina (g/dL) è stata calcolata mediante una formula matematica partendo dal valore di ematocrito: emoglobina (Hb, in g/dL) = ematocrito (Htc, in %) x 0,34.

Il gruppo sanguigno ABO e il fattore Rh sono stati testati mediante l'utilizzo di anticorpi monoclonali che, in caso di reazione positiva, provocavano emoagglutinazione.

Infine la glicemia (concentrazione di glucosio nel sangue, in mg/dL) è stata determinata solo per gli individui di Puno da personale del laboratorio dell'Hospital Regional di Puno.

Per la raccolta dei campioni di saliva è stato utilizzato il kit di auto-raccolta OG-250 distribuito da DNA Genotek (DNA Genotek Inc., 2 Beaverbrook Road, Kanata, Ontario, Canada K2K 1L1). I campioni raccolti con questo kit possono essere conservati a temperatura ambiente per anni. In appendice si trovano le istruzioni per la raccolta della saliva e il protocollo di estrazione del DNA (Appendice 1).

Il campione totale è costituito da 227 individui di cui 61 provenienti dal sito di campionamento di Cusco, 24 da quello di Cusibamba e 142 da Puno (Appendice 2).

I valori riassuntivi di alcune caratteristiche del campione sono riassunti in tabella 2.

	CUSCO (3326 m)		CUSIBAMBA (3640 m)		PUNO (3830 m)	
Numero individui	61		24		142	
Maschi, n (%)	33 (54)		7 (29)		40 (28)	
Femmine, n (%)	28 (46)		17 (71)		102 (72)	
	M	F	M	F	M	F
Età media, anni (d.s.)	26 (9)	22 (4)	43 (17)	42 (14)	39 (15)	32 (14)
Altezza media, cm (d.s.)	168,9 (5,5)	156,3 (5,9)	159,0 (5,5)	150,6 (5,3)	168,0 (5,6)	155,2 (6,0)
Peso medio, kg (d.s.)	63,0 (8,2)	51,2 (5,5)	57,6 (9,0)	51,8 (7,3)	74,5 (12,9)	61,5 (12,3)
BMI medio (d.s.)	22,1 (2,8)	21,0 (1,9)	22,7 (2,3)	22,8 (3,0)	26,4 (4,1)	25,6 (4,9)

Tab.2 Caratteristiche del campione.

BMI (Body Mass Index): Indice di Massa Corporea, dato dal rapporto tra il peso, espresso in Kg, ed il quadrato dell'altezza, espressa in metri. I valori normali sono compresi tra 18,5 e 24,9 secondo le tabelle dell' Organizzazione Mondiale della Sanità (http://apps.who.int/bmi/index.jsp?introPage=intro_3.html)

3.1.3 Risultati delle misurazioni fenotipiche

I valori medi delle misurazioni di alcune variabili misurate nel campione sono riassunti in tabella 3, dove sono indicati anche i valori di riferimento ottenuti dalla letteratura.

	CUSCO (3326 m s.l.m.)		CUSIBAMBA (3640 m s.l.m.)		PUNO (3830 m s.l.m.)		ANDINI (>3600 m s.l.m.)	BASSA QUOTA		
	M	F	M	F	M	F	M	M		
Pressione sanguigna sistolica (mmHg)	118,8 (12,1)	106,3 (10,3)	118,4 (14,3)	109,1 (12,9)	119,2 (14,5)	108,6 (13,1)	104,7 ^b (11,0)	127,0 ^a		
Pressione sanguigna diastolica (mmHg)	68,8 (10,5)	69,5 (8,7)	71,4 (6,9)	65,2 (7,8)	76,1 (11,1)	70,7 (10,3)	69,1 ^b (8,2)	76,0 ^a		
Battito cardiaco (bpm)	70,4 (9,5)	74,9 (7,3)	62,4 (5,8)	70,2 (6,8)	72,1 (14,1)	72,9 (9,3)	63,8 ^b (10,4)	70,0 ^a		
SpO ₂ (%)	94,4 (2,0)	92,2 (5,1)	90,2 (9,5)	90,3 (4,8)	93,5 (3,1)	91,6 (4,9)	92,2 ^a (2,8)	96,0 ^a		
Emoglobina (g/dl)	17,8 (1,2)	16,7 (1,7)	18,9 (1,6)	16,6 (1,2)	17,8 (2,0)	16,4 (1,5)	M	F	M	F
							17,5 ^b (1,93) 19,2 ^a	17,8 ^a	15,3 ^a (0,02)	13,4 ^a (0,02)
Ematocrito (%)	53,8 (3,8)	48,8 (5,5)	57,9 (5,6)	49,8 (3,7)	53,8 (6,2)	48,7 (4,7)	47,7 ^b (3,8)		M	F
									42-52	36-46
Glucosio (mg/dl)	-	-	-	-	69,1 (19,5)	67,6 (34,5)	50,6 ^c (3,7)	60-120		
Circonferenza del torace (cm) NORMALE	91,4 (6,4)	80,3 (4,8)	91,3 (3,1)	85,7 (6,1)	98,7 (9,5)	88,5 (11,1)	92,4 ^b (4,18)	93,3		
Circonferenza del torace (cm) MASSIMA	95,1 (6,6)	83,5 (4,9)	95,7 (3,3)	89,3 (6,1)	101,9 (9,5)	91,3 (10,7)	96,2 ^b (4,15)	99,0		
Circonferenza del torace (cm) MINIMA	87,5 (7,1)	77,4 (5,3)	88,7 (3,1)	83,4 (5,7)	96,6 (9,8)	86,4 (11,2)	90,4 ^b (4,18)	89,7		

Tab.3 Valori medi di variabili rilevate nel nostro studio confrontati con valori di riferimento presenti in letteratura per popolazioni andine e popolazioni nord americane di bassa quota. In parentesi sono indicati i valori di deviazione standard ove presenti.

^aBeall et al. 1997, Beall et al. 1998; Beall 2007b.

^bTarazona-Santos et al. 2000

^cCastillo et al. 2007

Di seguito gli andamenti di queste misure vengono commentati brevemente:

In figura 7 sono riportate le distribuzioni dei valori di pressione sistolica e diastolica in maschi e femmine del nostro campione. I valori medi possono essere discussi confrontandoli con i valori medi disponibili in letteratura per popolazioni andine e tibetane (Beall et al. 1997, Beall et al. 1998, Tarazona-Santos et al. 2000, Hoit et al. 2005, Beall 2007b, Beall 2007a) .

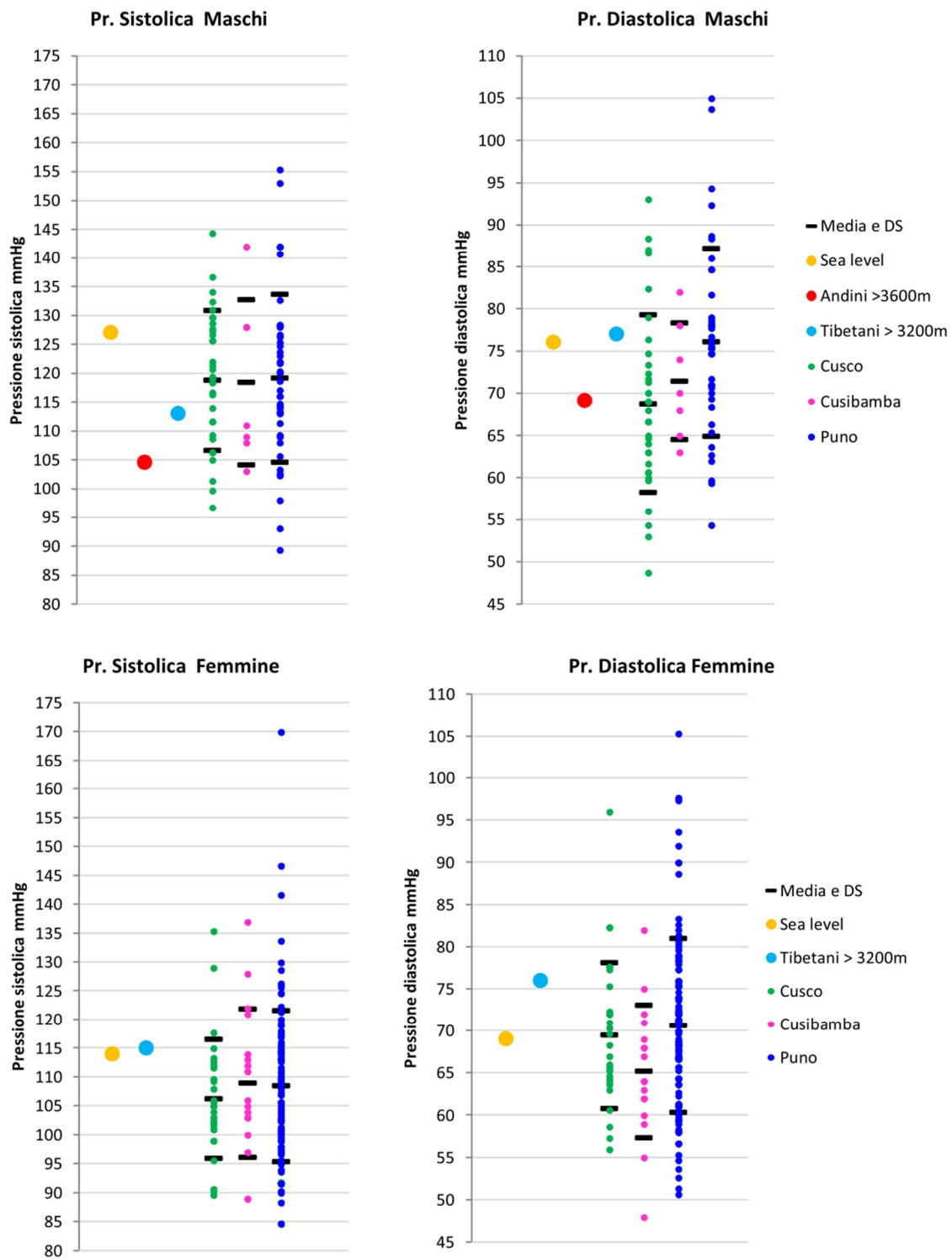


Fig.7 Rappresentazione grafica dei valori di pressione sanguigna sistolica e diastolica riferiti alla componente maschile e femminile delle popolazioni campionate in questo studio. Sono riportati nel grafico anche i valori medi di riferimento di bassa quota (Livello del mare) e dati della letteratura riferiti a popolazioni andine residenti al di sopra dei 3600m e tibetane residenti ad altitudini maggiori di 3200m. (Beall et al. 1997, Beall et al. 1998, Tarazona Santos et al. 2000, Hoit et al. 2005, Beall 2007b) Per le femmine non sono disponibili i valori di riferimento della letteratura relativi agli andini.

Dalla letteratura sappiamo che generalmente le popolazioni andine presentano valori di pressione sistolica e diastolica più bassi rispetto ai valori di riferimento per controlli di bassa quota e per popolazioni tibetane (Fiori et al. 2000).

Nei nostri campioni i valori medi di pressione sistolica non mostrano differenze significative nei tre siti di campionamento e concordano con i dati riscontrati in letteratura evidenziando valori inferiori rispetto alla media per popolazioni di bassa quota. Per quanto riguarda il confronto con il valore medio di popolazioni tibetane possiamo osservare come nel caso delle femmine i dati concordino con la letteratura indicando un valore medio superiore rispetto a quello delle popolazioni tibetane, mentre nei maschi tali valori sono paragonabili.

Per quanto riguarda la pressione diastolica si osservano delle differenze significative tra i valori medi misurati nei siti di Cusco e Puno per il campione maschile, tuttavia la situazione non è replicata nel campione femminile. Questo potrebbe dipendere dal fatto che gli individui maschi di Puno mostrano valori medi di indice di massa corporea (BMI) significativamente maggiori rispetto a quelli di Cusco (BMI medio Puno= 26, BMI medio a Cusco=22). Generalmente valori elevati di indice di massa corporea, superiori alla soglia di normalità (18,5-25), sono una delle cause di ipertensione. Tuttavia si può notare che il valore medio di pressione sistolica è paragonabile al valore medio di un individuo sano residente a bassa quota indicando assenza di ipertensione come atteso per le popolazioni andine.

È comunemente noto che mantenere valori bassi di pressione sanguigna riduce l'insorgenza di problemi cardiovascolari. Come descritto nei capitoli precedenti le popolazioni andine attivano un tipo di risposta alle basse pressioni di ossigeno che coinvolge principalmente il sistema circolatorio, favorendo un aumento della quantità di ossigeno trasportato grazie ad un incremento della concentrazioni di emoglobina e quindi di ematocrito. Tuttavia questo determina una maggiore esposizione a patologie del sistema cardiocircolatorio (Chronic Mountain Sickness, CMS) dovute ad una eccessiva densità del sangue, pertanto è possibile ipotizzare che il mantenimento di basse pressioni sanguigne sia un sistema per non sovraccaricare l'apparato circolatorio e ridurre il rischio, già elevato, di incorrere in tali patologie. A dimostrazione di questo, le popolazioni tibetane che attivano risposte che coinvolgono principalmente l'apparato respiratorio,

mantenendo bassi i livelli di emoglobina ed ematocrito, son esenti dal rischio di patologie come la CMS e presentano valori medi di pressione sanguigna maggiori rispetto alle popolazioni andine.

In figura 8 è possibile osservare i valori di emoglobina misurati nei nostri campioni e i relativi valori di riferimento della letteratura.

Rispetto a quanto appena descritto è interessante notare come anche i nostri campioni mostrino valori di emoglobina significativamente superiori rispetto ai valori medi misurati in individui nativi di bassa quota e in popolazioni tibetane. Tali valori sono paragonabili a quanto riscontrato in letteratura per popolazioni andine residenti a quote superiori ai 3600m (Tarazona-Santos et al. 2000, Beall 2007b).

In figura 8 sono riportati anche i valori di saturazione di ossigeno, ossia la percentuale di molecole di emoglobina saturate da molecole di ossigeno .

Tale valore dipende strettamente dalla quantità di ossigeno in circolo e rappresenta una misura diretta del livello di ossigenazione di un individuo. Quando la pressione di ossigeno nel sangue arterioso (pressione arteriosa) scende a valori inferiori a 60 mmHg la desaturazione accelera notevolmente secondo una curva esponenziale (curva di dissociazione dell'emoglobina). Il valore normale a livello del mare è compreso tra 95%-98% (Cogo 2009). I dati in letteratura mostrano che le popolazioni andine e tibetane hanno valori di saturazione di ossigeno mediamente più bassi di quelli riscontrati in popolazioni a livello del mare, tuttavia se confrontiamo le due popolazioni di alta quota gli andini mostrano valori significativamente maggiori rispetto alle popolazioni tibetane (84%) (Hoit et al. 2005, Beall 2007b). I nostri campioni andini concordano con quanto riscontrato in letteratura. Pertanto seppure questo fenotipo in popolazioni andine non sia associato a livelli significativi di ereditabilità (h^2) potrebbe comunque essere indice di un migliore adattamento poiché lo stress dovuto alla mancanza di ossigeno è ridotto rispetto a quello di popolazioni tibetane (Hoit et al. 2005).

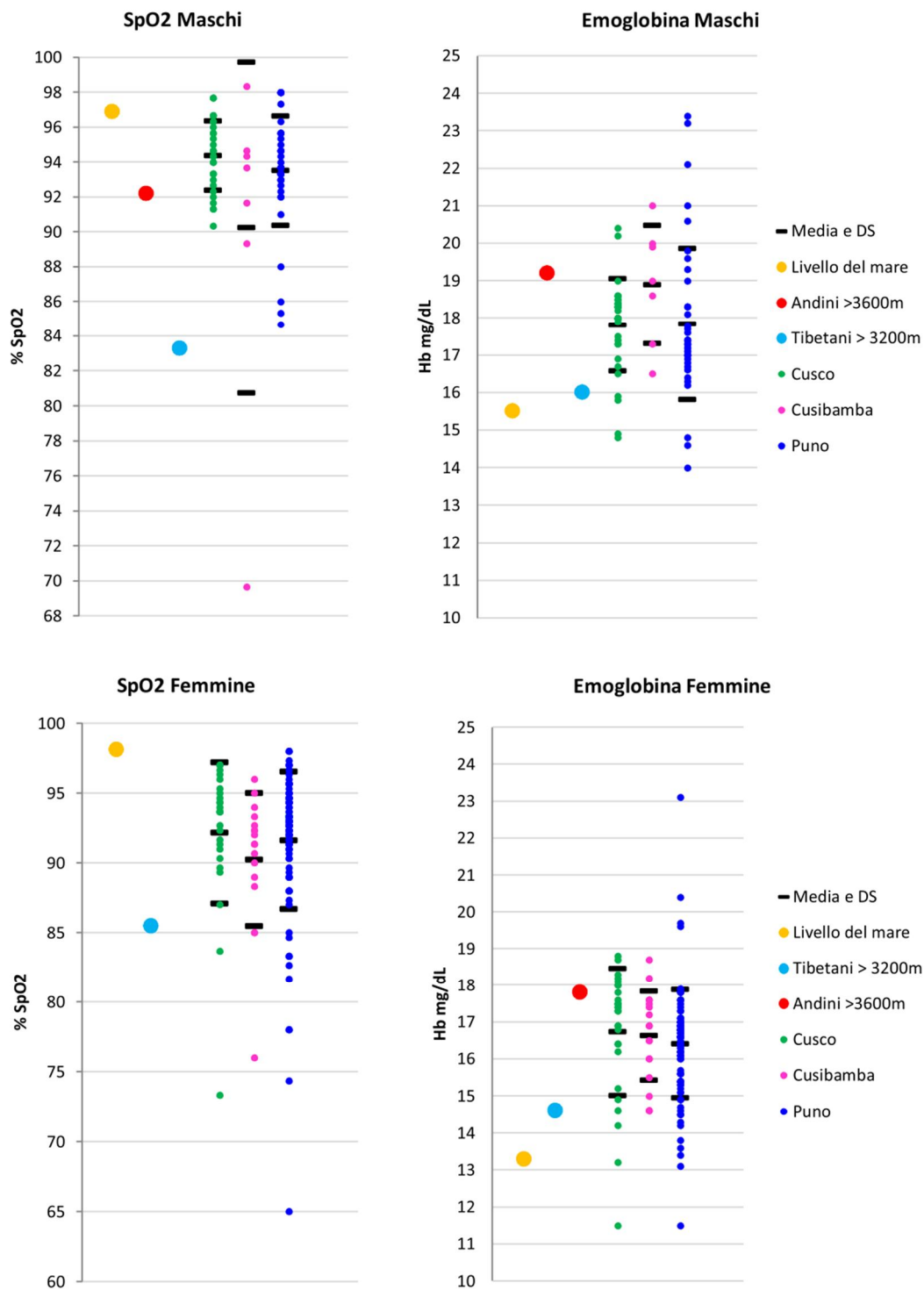


Fig.8 Rappresentazione grafica dei valori di saturazione di ossigeno (SpO₂) e concentrazione di emoglobina riferiti alla componente maschile e femminile delle popolazioni campionate in questo studio. Sono riportati nel grafico anche i valori medi di riferimento di bassa quota (Livello del mare) e dati della letteratura riferiti a popolazioni andine residenti al di sopra dei 3600m (non disponibili per il valore di SpO₂ nelle femmine) e tibetane residenti ad altitudini maggiori di 3200m. (Beall et al. 1997, Beall et al. 1998, Tarazona Santos et al. 2000, Fiori et al. 2000, Beall 2007b).

Per la popolazioni di Puno è stato possibile rilevare anche il valore di glicemia. In figura 9 sono riportati i valori misurati negli individui del nostro campione e i valori medi di riferimento per individui si bassa quota e per popolazioni andine della letteratura (Castillo et al. 2007). Accanto ai valori di glicemia sono riportati i valori di indice di massa corporea (BMI) strettamente correlato ai livelli di glicemia.

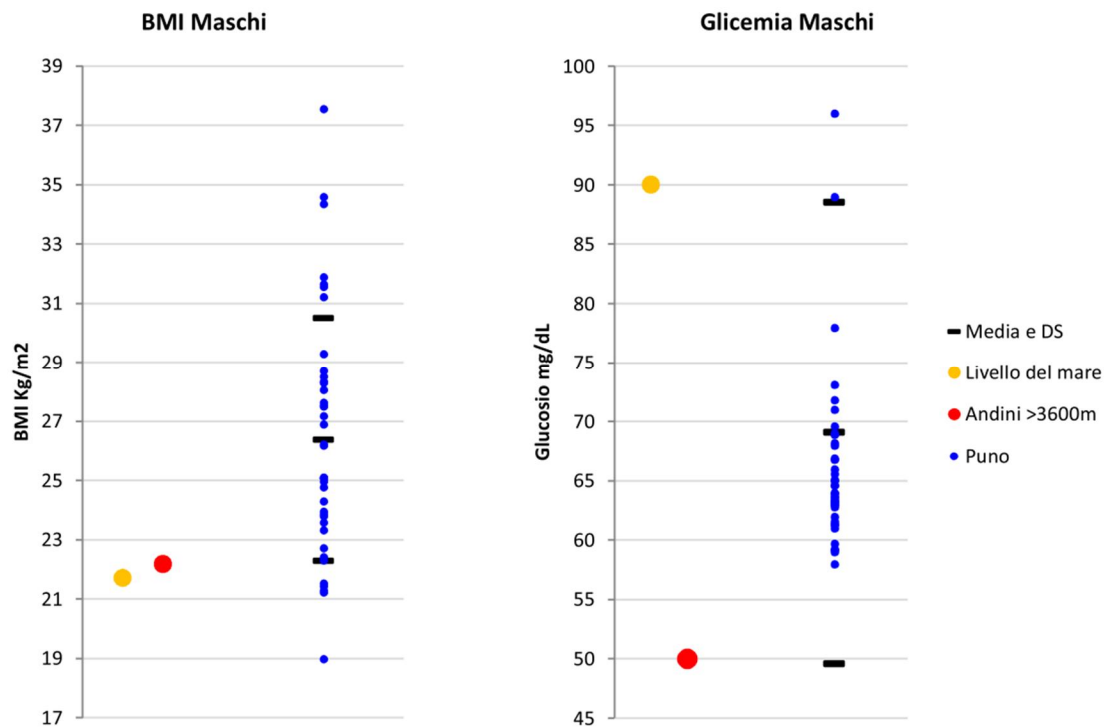


Fig.9 Rappresentazione grafica dei livelli di glucosio e di Indice di massa corporea (BMI) riferiti ad individui maschi di Puno. Sono riportati i valori medi di riferimento a bassa quota (livello del mare) e quelli disponibili in letteratura per popolazioni andine (Castillo et al. 2007, Beall et al. 1997, Beall 2007b; Fiori et al.2000, Tarazona-Santos et al. 2000).

È interessante notare come, nelle popolazioni andine di riferimento e nei nostri campioni di Puno, il valore di glicemia risulti significativamente più basso del valore medio a livello del mare. In particolare se si prende in considerazione il range di normalità (70-110 mg/dL) utilizzato in medicina per le diagnosi di patologie del metabolismo del glucosio, si può notare come a parità di BMI, i valori di glicemia negli andini si collocano al limite inferiore dell'intervallo di normalità, in accordo con quanto riscontrato in letteratura. Inoltre, dalla letteratura, sappiamo che questa riduzione nella concentrazione di glucosio non si associa a cambiamenti nei livelli di insulina, ma piuttosto si pensa sia imputabile ad

una inibizione di circa il 50% della gluconeogenesi (produzione di glucosio a partire dal glicogeno immagazzinato a livello epatico) a favore di una maggiore conversione del glucosio in glicogeno.(Castillo et al. 2007).

Al contrario in individui nativi di bassa quota che si recano in alta quota si osserva un incremento della glicemia dopo alcuni giorni in altitudine che poi torna a valori normali.

Attualmente non sono note spiegazioni di tipo adattativo, tuttavia è possibile ipotizzare che si tratti di un sistema di regolazione volto a ridurre il rischio di patologie, quali il diabete, in popolazioni la cui alimentazione è particolarmente ricca di carboidrati.

3.2 Il campione di confronto di bassa quota

In questo studio abbiamo confrontato la variabilità genetica osservata ad alcuni loci candidati con quella agli stessi loci osservata in individui nativi di bassa quota allo scopo di valutare se ci fossero differenze significative.

Idealmente il campione di confronto dovrà avere origine etnica simile a quella del campione di alta quota. In questo modo se la variabilità riscontrata nella popolazione di alta quota differisce significativamente da quella riscontrata in popolazioni provenienti dalla stessa area geografica, ma residenti a bassa quota, è possibile ipotizzare un effetto della pressione selettiva esercitata dall'ambiente.

La scelta del campione di bassa quota è stata quindi effettuata utilizzando un dataset di 687 microsatelliti autosomici tipizzati da Wang et al. (2007) in 5 popolazioni native americane (Maya e Pima provenienti dal Messico, Colombiani, Surui e Karitiana dal Brasile) provenienti dallo Human Genome Diversity Project (HGDP) (Cann et al. 2002), e una popolazione di individui Quechua campionati dagli autori in una località geografica vicina ai nostri siti di campionamento (Wang et al. 2007). Per verificare quale delle popolazioni native americane di bassa quota mostrasse minori differenze con le popolazioni andine di alta quota, sono state misurate le distanze genetiche tra le popolazioni utilizzando appunto i microsatelliti.

I marcatori microsatelliti (Short Tandem Repeats, STR) sono marcatori costituiti di unità di 1-5 bp ripetute in tandem dalle 10 alle 30 volte. Il polimorfismo è rappresentato dal numero di ripetizioni pertanto sono definiti marcatori multiallelici. Si trovano dispersi nell'intero genoma in regioni non codificanti e presentano un elevato tasso di mutazione (10^{-3} - 10^{-4} per locus per generazione) se paragonati con i tassi di mutazione puntiforme di loci genici codificanti (10^{-8}) (Jobling et al. 2004b).

I microsatelliti seguono un modello mutazionale di tipo "stepwise" secondo il quale la mutazione cambia la lunghezza dell' STR aggiungendo o sottraendo singole ripetizioni ad un tasso fisso che è indipendente dalla lunghezza dell'STR stesso (Ellegren 2004).

L'elevata variabilità di questi marcatori può aumentare il rischio di omoplasia. Per un marcatore genetico l'omoplasia si verifica quando, a causa di eventi mutazionali si hanno

copie diverse di uno stesso locus di ugual lunghezza, non identiche per discendenza. Inoltre quando i microsatelliti si presentano con dimensioni alleliche diverse il loro tasso di mutazione in un dato locus non è sempre lo stesso e non è costante nel tempo poiché il tasso di mutazione e la direzione della mutazione sono influenzati dalla lunghezza dell'allele (Ellegren 2000).

Nonostante ciò l'elevato polimorfismo e le caratteristiche di neutralità li rendono tra i marcatori più utilizzati per studi sulla struttura genetica delle popolazioni (Rosenberg et al. 2002, Ramachandran et al. 2005).

Il calcolo delle distanze genetiche tra coppie di popolazioni tramite l'indice F_{ST} (descritto nel paragrafo 4.2.4) è stato effettuato utilizzando il software Arlequin versione 3.5.1.2 (Excoffier and Lischer 2010). Questo indice può assumere valori compresi tra 0 e 1, dove 0 indica assenza di differenziazione fra le popolazioni, mentre valori crescenti indicano un progressivo aumento della distanza genetica tra le due popolazioni in esame.

Il confronto tra la popolazione Quechua e la popolazione Maya è quello che ha fornito il valore di F_{ST} più basso ($F_{ST} = 0,0098$) (Tab.4).

	Colombians	Karitina	Surui	Maya	Pima	Quechua
Colombians	0,00000					
Karitina	0,10754	0,00000				
Surui	0,15992	0,18989	0,00000			
Maya	0,04478	0,08035	0,12172	0,00000		
Pima	0,09925	0,13924	0,17160	0,05922	0,00000	
Quechua	0,04238	0,08128	0,12473	0,00985	0,06390	0,00000

Tab.4 F_{ST} a coppie calcolati utilizzando i dati microsatelliti di Wang et al.(2007).
Tutti i confronti sono significativi con p-value < 0,05

Pertanto il campione di controllo, costituito da individui nativi di bassa quota, utilizzato in questo studio è rappresentato da 25 individui Maya provenienti dallo Human Genome Diversity Project.

4. METODI

4.1 Protocolli di biologia molecolare

4.1.1 Estrazione del DNA

Il protocollo di estrazione del DNA è descritto in Appendice 1.

4.1.2 Amplificazioni mediante PCR

4.1.2.1 *mtDNA*

Amplificazione mediante PCR di 483 bp della regione ipervariabile HVRI:

Primer Light chain (L15931): 5'-ACCAGTCTTGTAACCGGAG-3'

Primer Heavy chain (H16414): 5'-CACGGAGGATGGTGGTCAAG-3'

La reazione è stata effettuata in un volume finale di 25 µL contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
10X PCR Buffer senza Mg	1 X
50 mM MgCl ₂	1 Mm
2 mM ogni dNTPs	0,16 mM ognuno
10µM Primer L15931	0,4 mM
10µM Primer H16414	0,4 mM
DNA campione	~ 20 ng
5U/µL Taq DNA Polymerase Recombinant, Invitrogen	1 U
Acqua distillata autoclavata	fino a 25 µl

Ciclo termico: 94°/2' + (94°/20'' + 52°/10'' + 70°/30'') x 30 cicli + 70°/5' + 10°/∞

Il prodotto della reazione è stato controllato su gel di agarosio all'1% .

Amplificazione mediante PCR di 95 bp della regione codificante *mtDNA*

Primer Light chain (L8209): 5'-CATCGTCCTAGAATTAATTCC-3'

Primer Heavy chain (H8304): 5'-CTTTACAGTGGGCTCTAGAGG-3'

La reazione è stata effettuata in un volume finale di 25 µl contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
10x PCR Buffer senza Mg	1 x
50 mM MgCl ₂	1 Mm
2 mM ogni dNTPs	0,16 mM ognuno
10µM Primer L8209	0,4 mM
10µM Primer H8304	0,4 mM
DNA campione	~ 20 ng
5U/µL Taq DNA Polymerase Recombinant, Invitrogen	1 U
Acqua distillata autoclavata	fino a 25 µl

Ciclo termico: 94°/2' + (94°/20'' + 57°/10'' + 70°/30'') x 30 cicli + 70°/5' + 10°/∞

La presenza/assenza della delezione è stata verificata mediante corsa elettroforetica su gel di agarosio al 3% .

4.1.2.2 EPO

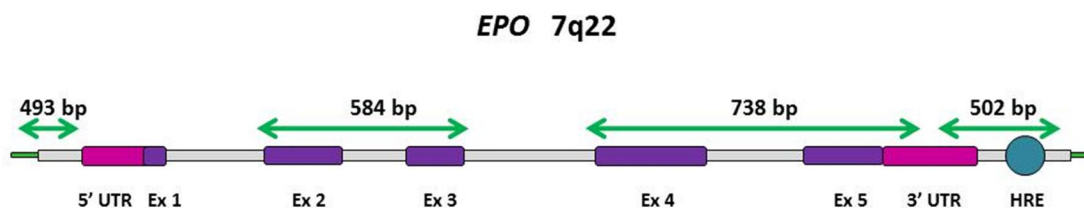


Fig.10 Struttura del gene *EPO*. I frammenti sequenziati sono indicati con frecce verdi. I box viola (Ex) indicano gli esoni, mentre i box fucsia si riferiscono alle estremità 5' e 3' non tradotte (UTR). Il cerchio azzurro rappresenta l'Hypoxia Responsive Element.

Amplificazione mediante PCR di 3 regioni del gene:

- Regione 1 : Promotore 586bp

Primer forward: EPO_PromF 5'-AAGGTCTTGCCATGTTGTCC-3'

Primer reverse: EPO_PromR 5'-GTTGCCATATTGCCCATACC-3'

- Regione 2 : esone 2 - esone3: 1004bp

Primer forward: EPO_Fs1 5'-GCAGCAGGATTGAATGAAGG-3'

Primer reverse: EPO_R1 5'-AGCTCACTGCAGCCTCAAAT-3'

- Regione 3 : esone 4 – esone5 – 3'UTR – HRE: 1578bp

Primer forward: EPO_F2 5'-TCTGTTTGCTCAGCTTGGTG-3'

Primer reverse: EPO_R2 5'-TAGCGAGGATGTGCATTGAG-3'

Regioni 1 e 2 : le reazioni sono state effettuate in un volume finale di 25 μ l contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
10x PCR Buffer senza Mg	1 x
50 mM MgCl ₂	1Mm
2 mM ogni dNTPs	0,16 mM ognuno
10 μ M Primer Forward	0,4 mM
10 μ M Primer Reverse	0,4 mM
DNA campione	~ 20 ng
5U/ μ L Taq DNA Polymerase Recombinant, Invitrogen	1 U
Acqua distillata autoclavata	fino a 25 μ l

Regione 1 :

Ciclo termico: $94^{\circ}/2' + (94^{\circ}/20'' + 57^{\circ}/10'' + 72^{\circ}/30')$ x 30 cicli + $72^{\circ}/5' + 10^{\circ}/\infty$

Regione 2 :

Ciclo termico: $94^{\circ}/2' + (94^{\circ}/20'' + 56^{\circ}/10'' + 72^{\circ}/1')$ x 30 cicli + $72^{\circ}/5' + 10^{\circ}/\infty$

Regione 3: le reazioni sono state effettuate in un volume finale di 50 μ l contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
3,3 x PCR Buffer XL	1 x
25 mM Mg(OAc) ₂	1,25Mm
2 mM ogni dNTPs	0,2mM ognuno
10 μ M Primer Forward	0,4 mM
10 μ M Primer Reverse	0,4 mM
DNA campione	~ 20 ng
2U/ μ L Taq GeneAmp XL	1 U
Acqua distillata autoclavata	fino a 50 μ l

Ciclo termico:

$94^{\circ}/1' + (94^{\circ}/30'' + 60^{\circ}/10')$ x 9 cicli + $(94^{\circ}/30'' + 60^{\circ}/10' + 15''_{+15''\text{per ciclo}})$ x 24 cicli + $72^{\circ}/30' + 10^{\circ}/\infty$

I prodotti delle reazioni sono stati controllati su gel di agarosio all'1% .

4.1.2.3 VHL



Fig.11 Struttura del gene *VHL*. I frammenti sequenziati sono indicati con frecce verdi. I box azzurri (Ex) indicano gli esoni, mentre i box blu si riferiscono alle estremità 5' e 3' non tradotte (UTR).

Amplificazione mediante PCR di 3 regioni del gene:

- Regione 1 : 5'UTR ed esone1: 1076bp

Primer forward: VHL Ex1F 5'-AAGCGTGATGATTGGGTGTT-3'

Primer reverse: VHL Ex1R 5'-GCTTCAGACCGTGCTATCGT-3'

- Regione 2 : esone 2 e regioni fiancheggianti: 1206bp

Primer forward: VHL_Ex2_fwd 5'-GTGCCCTGACTCAGACCAGT-3'

Primer reverse: VHL_Ex2_rev 5'-GATTGGCACTGTGAGCTTGA-3'

- Regione 3 : esone 3 e parte della regione 3'UTR: 1828bp

Primer forward : VHL_Ex3_fwd 5'-TCTAGCAGGGTAGCTCAGGG-3'

Primer reverse : VHL_Ex3_rev 5'-CCAGGCTACTCCATGAATCC-3'

Le reazioni sono state effettuate in un volume finale di 25 μ l contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
10x PCR Buffer senza Mg	1 x
50 mM MgCl ₂	1/1,5 Mm
2 mM ogni dNTPs	0,16 mM ognuno
10 μ M Primer Forward	0,4 mM
10 μ M Primer Reverse	0,4 mM
DNA campione	~ 20 ng
5U/ μ L Taq DNA Polymerase Recombinant, Invitrogen	1 U
Acqua distillata autoclavata	fino a 25 μ l

Ciclo termico: $94^{\circ}/2' + (94^{\circ}/20'' + 57^{\circ}/10'' + 72^{\circ}/1') \times 30$ cicli + $72^{\circ}/5' + 10^{\circ}/\infty$

Il prodotto della reazione è stato controllato su gel di agarosio all'1% .

4.1.2.4 EPAS1

Amplificazione mediante PCR di 2 regioni del gene:

- Regione 1 : 1208bp (842bp dell'introne 4, tutto l'esone 5 e 119bp dell'esone 6)

Primer forward: EPAS1 Ex5-6F 5'-GATGGTGGTGTAGGCAGGAT-3'

Primer reverse: EPAS1 Ex5-6R 5'-TGCTGGATTGGTTCACACAT-3'

- Regione 2 : 1056bp (regione tra l'esone 8 e l'esone 9 inclusi)

Primer forward: EPAS1 Ex8-9F 5'-TGCACCAAGGGTCAGGTAGT-3'

Primer reverse: EPAS1 Ex8-9R 5'-CCAGAGAGATGATGGCGTCT-3'

Le reazioni sono state effettuate in un volume finale di 25 μ l contenente:

<i>Reagenti</i>	<i>Concentrazioni finali</i>
10x PCR Buffer senza Mg	1 x
50 mM MgCl ₂	1,5 Mm
2 mM ogni dNTPs	0,16 mM ognuno
10 μ M Primer Forward	0,4 mM
10 μ M Primer Reverse	0,4 mM
DNA campione	~ 20 ng
5U/ μ L Taq DNA Polymerase Recombinant, Invitrogen	1 U
Acqua distillata autoclavata	fino a 25 μ l

Ciclo termico: $94^{\circ}/2' + (94^{\circ}/20'' + 57^{\circ}/10'' + 72^{\circ}/1') \times 30$ cicli + $72^{\circ}/5' + 10^{\circ}/\infty$

Il prodotto della reazione è stato controllato su gel di agarosio all'1% .

I prodotti delle reazioni di PCR sono stati purificati per via enzimatica allo scopo di eliminare i residui di Primer e dNTPs utilizzando:

- 5 unità di Exonuclease I (ExoI)
- 1 unità di Shrimp Alkaline Phosphatase (SAP) (USB Corporation)
- 5 µl di prodotto di PCR

Profilo termico: 37°/15' + 80°/15'

4.1.3 Sequenziamento

4.1.3.1 *mtDNA*

Il sequenziamento della regione amplificata ha portato ad ottenere una sequenza di 359 bp compresa tra la posizione 16024 e la posizione 16383.

Il primer utilizzato per il sequenziamento è L15931 utilizzato per la reazione di PCR.

La reazione di sequenziamento è stata effettuata in un volume finale di 10 µl:

<i>Reagenti</i>	<i>Quantità finali</i>
BigDye® Terminator v3.1 Cycle Sequencing RR-100 Mix kit, Applied Biosystems.	1 µl
BigDye® Terminator v1.1, v3.1 2.5X Buffer di sequenziamento	3 µl
3.2 µM Primer L15931	1 µl
DNA purificato	2 µl
Acqua distillata autoclavata	fino a 10 µl

Ciclo termico: (96°/10'' + 55°/5'' + 60°/1'30'') x 25 cicli + 10°/ ∞

4.1.3.2 EPO

Primer utilizzati per il sequenziamento delle 3 regioni del gene amplificate:

- Regione 1: abbiamo ottenuto un frammento di 493bp

Primer forward: EPO_PromF 5'-AAGGTCTTGCCATGTTGTCC-3'

- Regione 2: abbiamo ottenuto un frammento di 584bp

Primer forward: EPO_Fs1 5'-GCAGCAGGATTGAATGAAGG-3'

- Regione 3 : è stata sequenziata in due frammenti di 738bp e 502bp per un totale di 1240bp. I due frammenti non sono sovrapposti per 223bp.

Primer forward: EPO_F2 5'-TCTGTTTGCTCAGCTTGGTG-3'

Primer reverse: EPO_R2 5'-TAGCGAGGATGTGCATTGAG-3'

La reazione di sequenziamento è stata effettuata in un volume finale di 10 µl:

<i>Reagenti</i>	<i>Quantità finali</i>
BigDye® Terminator v3.1 Cycle Sequencing RR-100 Mix kit, Applied Biosystems.	1 µl
BigDye® Terminator v1.1, v3.1 2.5X Buffer di sequenziamento	3 µl
3.2 µM Primer	1 µl
DNA purificato	2 µl
Acqua distillata autoclavata	fino a 10 µl

Ciclo termico: (96°/10'' + 57°/5'' + 60°/1'30'') x 25 cicli + 10°/ ∞

4.1.3.3 VHL

Primer utilizzati per il sequenziamento delle 3 regioni del gene amplificate:

- Regione 1: abbiamo ottenuto un frammento di 635bp

Primer reverse: VHL Ex1R 5'-GCTTCAGACCGTGCTATCGT-3'

- Regione 2: abbiamo ottenuto un frammento di 319bp

Primer reverse: VHL_Ex2_Seq1 5'-CTGTACTIONACCACAACAACC-3'

- Regione 3 : è stata sequenziata in due frammenti di 685bp e 571bp per un totale di 1256bp. I due frammenti non sono sovrapposti per 217bp.

Primer forward: VHL_Ex3_FSeq1 5'-CCTTGTACTIONGAGACCCTAG-3'

Primer forward: VHL_Ex3_FSeq2 5'-GTGCACTIONTTCTCGGTCCACT-3'

La reazione di sequenziamento è stata effettuata in un volume finale di 10 µl:

<i>Reagenti</i>	<i>Quantità finali</i>
BigDye® Terminator v3.1 Cycle Sequencing RR-100 Mix kit, Applied Biosystems.	1 µl
BigDye® Terminator v1.1, v3.1 2.5X Buffer di sequenziamento	3 µl
3.2 µM Primer	1 µl
DNA purificato	2 µl
Acqua distillata autoclavata	fino a 10 µl

Ciclo termico: (96°/10'' + 57°/5'' + 60°/1'30'') x 25 cicli + 10°/ ∞

4.1.3.4 EPAS1

Primer utilizzati per il sequenziamento delle 2 regioni del gene amplificate:

- Regione 1: abbiamo ottenuto un frammento di 972bp sequenziato con due primer

Primer forward: EPAS1 Ex5-6Fs 5'-GGAGGCTGTTTGGCAACTAT-3'

Primer reverse: EPAS1 Ex5-6Rs 5'-GTGAGGAGGGCAGTTGTTGT-3'

- Regione 2: abbiamo ottenuto un frammento di 828bp sequenziato con due primer

Primer forward: EPAS1 Ex8-8Fs 5'-CATCTACAACCCTCGCAACC-3'

Primer reverse: EPAS1 Ex8-9Rs 5'-GGCTCCTCCTTTAGCTTGGT-3'

La reazione di sequenziamento è stata effettuata in un volume finale di 10 µl:

<i>Reagenti</i>	<i>Quantità finali</i>
BigDye® Terminator v3.1 Cycle Sequencing RR-100 Mix kit, Applied Biosystems.	1 µl
BigDye® Terminator v1.1, v3.1 2.5X Buffer di sequenziamento	3 µl
3.2 µM Primer	1 µl
DNA purificato	2 µl
Acqua distillata autoclavata	fino a 10 µl

Ciclo termico: (96°/10'' + 57°/5'' + 60°/1'30'') x 25 cicli + 10°/ ∞

I prodotti delle reazioni di sequenziamento sono stati purificati tramite precipitazione con etanolo secondo il seguente protocollo:

- Preparare una mix contenente 60 μ l di EtOH assoluto, 3 μ l di NaAc 3M e 27 μ l di H₂O.
- Aggiungere 90 μ l di mix ad ogni campione e lasciare 15 minuti a temperatura ambiente, invertendo un po' di volte.
- Centrifugare 30 minuti a 3000 X g.
- Liberarsi dell'EtOH invertendo i tubi con uno o più colpi decisi.
- Aggiungere 150 μ l di EtOH al 70% conservato a -20°C.
- Agitare invertendo un po' di volte.
- Centrifugare a 3000 x g per 10 minuti.
- Liberarsi dell'EtOH invertendo i tubi con uno o più colpi decisi.
- Centrifugare i campioni invertiti per 1 minuto a 1000 X g.

La corsa elettroforetica su sequenziatore automatico 3730xl e 3130xl è stata effettuata presso BMR Genomics Srl o presso StarSEQ[®] GmbH.

4.2 Analisi statistiche

4.2.1 Determinazione degli aplotipi

Le tecniche di genotipizzazione utilizzate in questa tesi, forniscono informazioni circa il genotipo dell'individuo, ma non permettono di determinare la fase cromosomica, cioè come le mutazioni sono associate tra loro a formare un aplotipo.

La fase può essere determinata con metodi di laboratorio come il clonaggio, oppure inferita con tecniche statistiche a partire dai genotipi.

A questo scopo è stato utilizzato l'algoritmo Expectation Maximization, EM (Excoffier and Slatkin 1995) contenuto nel software ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010).

L'algoritmo si basa su un processo iterativo mirato ad ottenere una stima della massima likelihood delle frequenze aplotipiche a partire da dati genotipici multi-locus per i quali la fase gametica è ignota. Trattandosi in questo caso di individui che possono presentare eterozigosi multipla non è possibile effettuare un semplice calcolo manuale poiché ci sono diversi genotipi possibili.

La likelihood del campione (la probabilità di osservare i dati D , date le frequenze aplotipiche p) si esprime come:

$$L(D|p) = \sum_{i=1}^n \prod_{j=1}^{g_i} G_{ij}$$

Dove n è il numero di individui del campione, e G_{ij} rappresentano tutti i possibili genotipi del campione, $G_{ij}=2p_i p_j$ se $i \neq j$, $G_{ij}=p_i^2$ se $i=j$.

Il procedimento su cui si basa l'algoritmo EM è il seguente:

1. Inizia con una stima arbitraria delle frequenze aplotipiche.
2. Utilizza questa stima per calcolare le frequenze attese dei genotipi assumendo l'Equilibrio di Hardy-Weinberg (fase E).
3. Le frequenze relative dei genotipi vengono poi utilizzate per calcolare gli aplotipi ad essi riferiti mediante una procedura di conteggio che porta alla stima di nuove frequenze aplotipiche (fase M).
4. Si ripetono gli step 2 e 3 finché le frequenze non raggiungono l'equilibrio, ossia non variano più di un prefissato valore epsilon.

Per fare questo calcolo l'algoritmo EM prevede come assunzione che sia rispettata la condizione di equilibrio di Hardy-Weinberg.

Trattandosi di una stima di Massima likelihood è possibile che le frequenze aplotipiche che massimizzano il valore di likelihood dei dati siano dovute al raggiungimento di un massimo locale durante l'esplorazione dello spazio dei parametri. Questo determina una interruzione dell'iterazione e il valore ottenuto non corrisponde a quello che si otterrebbe esplorando lo spazio globale.

Per ovviare a questo problema si deve ripetere l'algoritmo EM a partire da diversi punti dello spazio dei parametri (in questo caso diverse frequenze aplotipiche iniziali) e infine scegliere il valore di frequenza che presenta il valore di likelihood maggiore.

La deviazione standard delle frequenze aplotipiche viene stimata mediante una procedura di ri-campionamento (bootstrap) in cui vengono generati campioni casuali da una popolazione assumendo che abbiano frequenze aplotipiche uguali al corrispondente valore che massimizza la likelihood. Per ogni replica viene applicato l'algoritmo EM e la deviazione standard di ogni frequenza aplotipica viene stimata dalla distribuzione delle frequenze aplotipiche che ne risulta (Excoffier and Slatkin 1995).

In questo studio gli aplotipi sono stati stimati con l'algoritmo EM impostando 100 come numero di punti di partenza per l'iterazione e 1000 come numero massimo di iterazioni. Il numero di bootstrap è stato impostato a 1000.

4.2.2 Stima del linkage disequilibrium

Quando due o più marcatori genetici sono trasmessi insieme con una frequenza maggiore di quella che ci si attende per caso sulla base delle loro frequenze, si parla di associazione allelica. La loro cosegregazione, tuttavia, può essere dovuta a meccanismi che non necessariamente implicano associazione fisica sullo stesso cromosoma, come per esempio la mescolanza incompleta fra due sottopopolazioni (Barbujani e Goldstein 2004). Se, al contrario, i due alleli si trovano legati fisicamente sullo stesso cromosoma e per questo ereditati insieme con una frequenza maggiore di quella attesa se la segregazione fosse casuale, si parla di linkage disequilibrium (LD). Maggiore è la distanza fra due loci sul cromosoma, maggiore è la probabilità che la ricombinazione, omologa o non reciproca (conversione genica), causi il riassortimento degli alleli, quindi la diminuzione di LD fra questi ultimi.

Misurare il livello di linkage disequilibrium permette dunque di inferire l'entità degli eventi di ricombinazione genica ed il ruolo di questi ultimi nel modellare la diversità genetica osservata in una specifica regione del genoma.

Il livello di LD è stato misurato fra ogni coppia di posizioni polimorfiche in ogni campione tipizzato in questa tesi.

La misura classica di linkage disequilibrium D si esprime come:

$$D_{ij} = p_{ij} - p_i p_j$$

dove p_{ij} è la frequenza dell'aplotipo che presenta l'allele i al primo locus e l'allele j al secondo, mentre p_i e p_j sono le frequenze degli alleli i e j , rispettivamente. Se non c'è associazione, ci si aspetta che il valore di p_{ij} coincida col prodotto di p_i e p_j (Lewontin and Kojima 1960).

Il valore di D_{ij} risente perciò delle frequenze alleliche, e varia tra un minimo ed un massimo non definiti a priori.

In questo lavoro è stato utilizzato il coefficiente D_{ij} standardizzato per il massimo valore che può assumere ($D_{ij \max}$), date le frequenze alleliche (Lewontin 1964). Questo indice, detto D' , si esprime come:

$D'_{ij} = D_{ij} / D_{ij \max}$, dove $D_{ij \max}$ ha uno dei seguenti valori:

$\min(p_i p_j, (1-p_i)(1-p_j))$ se $D_{ij} < 0$

$\min((1-p_i)p_j, p_i(1-p_j))$ se $D_{ij} > 0$

Questo indice è indipendente dalle frequenze alleliche, essendo definito come la proporzione del massimo valore di D e varia fra +1 e -1. Il segno è irrilevante, in quanto dipende dalla scelta (arbitraria) degli alleli i e j rispetto alla loro alternativa.

$D' = 0$ indica assenza di associazione fra i loci considerati (indipendenza).

Un'altra misura proposta per valutare il livello di LD tra coppie di alleli a due loci è rappresentata dall'indice R^2 , ossia il quadrato del coefficiente di correlazione tra le frequenze alleliche espresso come funzione della misura di linkage disequilibrium D :

$$R^2 = \frac{D_{ij}^2}{p_i(1-p_i)p_j(1-p_j)}$$

$R^2=1$ si può ottenere solo quando le frequenze degli alleli i e j , p_i e p_j , sono uguali.

Per questa analisi è stato utilizzato il software DnaSP 5.10 (Librado and Rozas 2009). Il software calcola la significatività del test utilizzando un test esatto di Fisher a due code (Sokal and Rohlf 1981).

4.2.3 Variabilità genetica intra – popolazione

4.2.3.1 Indici di diversità

Per indagare la variabilità genetica all'interno di ciascuna popolazione abbiamo utilizzato diversi parametri:

- Numero di siti polimorfici (S): numero di siti segreganti nel campione considerato;
- Numero di aplotipi (k): numero di sequenze diverse nel nostro campione;
- Diversità aplotipica (H): probabilità che due aplotipi, scelti a caso nel campione, siano diversi ; viene calcolata come $\hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$ dove n è il numero di copie del gene nel campione, k è il numero di aplotipi e p_i è la frequenza dell'aplotipo i-esimo nel campione (Nei 1987).

Abbiamo inoltre utilizzato differenti stimatori del parametro mutazionale nella popolazione definito come $\theta = 2nN_e\mu$ dove μ è il tasso di mutazione per sito per generazione, N_e la dimensione effettiva della popolazione e n sarà uguale a "2" per loci diploidi, "1,5" per il cromosoma X e "0,5" per l'mtDNA e il cromosoma Y.

θ è il valore all'equilibrio della diversità genetica in un modello mutazione – deriva che assume: popolazione costante, assenza di migrazione e selezione, accoppiamento casuale (Jobling et al. 2004a).

Gli stimatori di θ che abbiamo utilizzato sono i seguenti:

- θ_s : basato sul numero dei siti polimorfici (S) (Watterson 1975)
- θ_π : basato sul numero medio di differenze a coppie tra sequenze (Tajima 1983);

Un altro indice stimato è π , definito anche "*diversità nucleotidica*", ed è dato dal numero medio di differenze a coppie tra sequenze per sito analizzato (Nei 1987).

Le distanze molecolari tra aplotipi sono state calcolate considerando il modello di sostituzione di Tamura e Nei (Tamura and Nei 1993). Questo modello permette di correggere il semplice calcolo del numero di nucleotidi per cui due sequenze differiscono

considerando diversi tassi di mutazione per le trasversioni e per le transizioni e diversi tassi mutazionali delle transizioni se avvengono tra pirimidine o tra purine. Nel caso del DNA mitocondriale è stata considerata anche l'eterogeneità del tasso di mutazione lungo la sequenza.

Il valore generalmente considerato nel caso del DNA mitocondriale per operare una correzione che consideri questa eterogeneità nel calcolare le distanze fra sequenze è 0,26 (Meyer et al. 1999).

4.2.3.2 Test di neutralità

I test di neutralità effettuati per il mtDNA sono: il D di Tajima e l' F_s di Fu.

- D di Tajima: $D = \frac{\theta_{\pi} - \theta_s}{SE(\theta_{\pi} - \theta_s)}$

dove θ_{π} è il numero medio di differenze a coppie tra sequenze, θ_s si basa sul numero di siti polimorfici nel campione (segreganti, S) $\theta_s = \frac{S}{\sum_{i=1}^{n-1} (1/i)}$ e SE è

l'errore standard (Tajima 1989).

In condizioni di neutralità (assenza di selezione e di ricombinazione, assenza di variazioni nelle dimensioni delle popolazioni o suddivisione delle stesse, equilibrio mutazione-deriva) le stime di theta (θ_{π} e θ_s) assumono lo stesso valore, pertanto la statistica D è uguale a zero.

Quando le assunzioni del modello neutrale vengono violate le stime di questi parametri subiscono delle variazioni. Ad esempio il parametro θ_s è fortemente influenzato da fenomeni quali espansioni delle popolazioni o selective sweep che causano un aumento della frequenza di alleli rari nelle popolazioni e quindi un aumento del numero di siti polimorfici (S). Tali eventi non determinano variazioni significative nel numero medio di differenze tra coppie di sequenze (θ_{π}) e pertanto il test D subirà deviazioni significative verso valori negativi.

Viceversa in condizioni di suddivisione delle popolazioni o fenomeni di selezione bilanciante viene favorito il mantenimento di varianti polimorfiche a frequenze intermedie influenzando principalmente le stime di θ_{π} . In questi casi si avranno

deviazioni significative verso valori positivi di questo test (Tajima 1989, Przeworski et al. 2000, Nielsen 2001, Yang 2006).

- F_s di F_u : questo test si basa sulla probabilità di osservare un campione neutrale casuale con un numero di alleli simile o minore del valore realmente osservato dato il numero di differenze a coppie espresso come stima di θ ; questa probabilità è stata definita come: $S' = \Pr(K \geq k_{obs} | \theta = \theta_\pi)$ e quindi $F_s = \ln\left(\frac{S'}{1-S'}\right)$.

Questo test è molto sensibile alle espansioni demografiche poiché si basa principalmente sul numero di siti polimorfici (S) che come detto precedentemente è fortemente influenzato da fenomeni di espansione delle popolazioni. Ciò determina valori negativi molto grandi dell' F_s (Fu 1997).

Per il locus *VHL* sono stati effettuati i test di neutralità D di Tajima descritto sopra e la D di Fu&Li (Fu and Li 1993).

- D di Fu&Li: questo test prende in considerazione la distribuzione delle mutazioni nella genealogia di un campione casuale di geni in una popolazione. Confrontando il numero di mutazioni nei rami interni ed esterni della genealogia con quanto atteso secondo le ipotesi di neutralità, sarà possibile rilevare gli effetti della selezione.

Pertanto la statistica D si basa sulla differenza tra η_e , il numero totale di mutazioni sui rami esterni di una genealogia (singletons), e η_i , il numero di mutazioni nei rami interni:

$$D = \frac{\eta_i - (a_n - 1)\eta_e}{SE(\eta_i - (a_n - 1)\eta_e)}$$

Dove: $a_n = \sum_{i=1}^{n-1} 1/i$, SE è l'errore standard e n il numero di sequenze (Fu and Li 1993, Yang 2006).

Secondo Fu e Li (1993) mutazioni deleterie tenderanno ad essere selezionate in modo negativo, per cui a presentarsi come singletons (mutazioni presenti una sola

volta) sui rami esterni della genealogia contribuendo ad aumentare η_e e indirizzare il test verso valori negativi.

Viceversa mutazioni nei rami interni della genealogia determineranno un incremento dell'indice η_i con conseguenti valori positivi del test.

Questo test è stato effettuato utilizzando DnaSP 5.10 (Librado and Rozas 2009) e utilizzando come sequenza outgroup la sequenza dello scimpanzé.

Assumendo il modello a siti infiniti (secondo cui ogni evento mutazionale colpisce un sito diverso) DnaSP calcola il numero totale di mutazioni nei rami esterni della genealogia come la somma per ogni sito polimorfico del numero di singleton che non sono condivisi con l'outgroup. La significatività del test viene calcolata utilizzando come valori critici quelli esposti nelle tabelle 2 e 4 del lavoro di Fu and Li (1993).

4.2.4 Variabilità genetica tra popolazioni

Per misurare la distanza genetica tra coppie di popolazioni abbiamo utilizzato l'indice F_{ST} molecolare.

Questo indice è un analogo dell' F_{ST} di Wright (Wright 1950) basato sull'informazione ottenuta dalla stima delle differenze fra coppie di aplotipi. Questo indice può assumere valori compresi in un intervallo tra 0 e 1, dove 0 indica assenza di differenziazione fra popolazioni, mentre valori crescenti indicano un progressivo aumento della distanza genetica tra le due popolazioni in esame. Per calcolare il p-value del test viene creata una distribuzione nulla dei valori di F_{ST} a coppie sotto l'ipotesi di assenza di differenziazione tra le popolazioni, mediante processi di permutazione. Il p-value del test è dato dalla proporzione di permutazioni che permette di ottenere il valore di F_{ST} maggiore o uguale a quello osservato. I valori di significatività vengono anch'essi mostrati sotto forma di matrice.

Per rappresentare graficamente la matrice di distanze genetiche abbiamo utilizzato il Multidimensional Scaling non parametrico (NMMDS). Questa analisi statistica trasforma le distanze genetiche in una matrice di (n x n) ranghi e le converte in distanze che possono essere rappresentate in uno spazio euclideo n-dimensionale. Il numero di dimensioni viene progressivamente ridotto attraverso un processo iterativo fino a poter rappresentare graficamente le distanze in due dimensioni. La bontà dell'adattamento della rappresentazione bidimensionale viene valutata attraverso un indice chiamato *stress* che è proporzionale alla distorsione generata dal processo iterativo. Per l'analisi MDS è stato utilizzato il programma STATISTICA versione 7.1 (StatSoft).

4.2.5 Test di associazione

In questa tesi sono stati utilizzati alcuni test di statistica parametrica e non parametrica per verificare le associazioni tra i genotipi ai loci *EPO* e *VHL* e i fenotipi misurati all'interno del campione di alta quota. Inoltre per il gene *VHL* abbiamo verificato se ci fosse associazione tra un dato genotipo e l'appartenenza al campione di alta quota (Andini) o di bassa quota (Maya).

4.2.5.1 Il test esatto di Fisher

Questo test permette di confrontare due o più serie di dati per verificare l'ipotesi nulla di indipendenza tra le variabili. Si basa sulla costruzione di tabelle di contingenza e viene utilizzato nel caso in cui più 20% delle osservazioni presentano una frequenza attesa inferiore a 5 o sono presenti celle con frequenza attesa minore di 1 (Whitlock and Schluter 2010), condizioni per cui non è possibile applicare il test del χ^2 .

Attraverso il calcolo combinatorio vengono calcolate e sommate le probabilità corrispondenti alla tabella osservata e quelle di tutte le tabelle osservabili che si discostano maggiormente, rispetto a quella osservata, dall'ipotesi di indipendenza, aventi gli stessi totali marginali della tabella osservata. Il calcolo computazionale è piuttosto complesso e pertanto si ricorre a software statistici in grado di automatizzare l'analisi.

Il test di Fisher è noto per tabelle di contingenza 2x2, tuttavia esistono estensioni del test anche per tabelle di contingenza 2xn gruppi. Per queste analisi abbiamo utilizzato il sito web "VassarStats: Website for Statistical Computation" che permette di effettuare il test automaticamente fornendo al programma i valori delle singole celle (<http://faculty.vassar.edu/lowry/webtext.html>).

Mediante questo test è stato verificato se i genotipi ai loci *EPO* e *VHL* si distribuiscono in modo indipendente nelle 2 categorie rappresentate dal gruppo etnico (Quechua e Aymarà) allo scopo di evitare l'attribuzione di una associazione genotipo/fenotipo erranea (stratificazione), dovuto a questa variabile.

Il test è stato utilizzato anche per verificare l'associazione tra un determinato genotipo al locus *VHL* e l'appartenenza al gruppo di alta e bassa quota, e per effettuare l'analisi dei fenotipi estremi.

L'analisi dei fenotipi estremi consiste nel confrontare la distribuzione della variabilità genetica ai loci *EPO* e *VHL*, tra gruppi di individui di alta quota con valori fenotipici al di sotto e al di sopra di una determinata soglia, allo scopo di massimizzare la probabilità di identificare una associazione tra genotipo e fenotipo (campionamento ottimale di fenotipi estremi).

4.2.5.2 ANOVA (ANalysis Of VAriance)

L'analisi della varianza è un metodo statistico che permette di confrontare simultaneamente le medie di più gruppi in una singola analisi.

Si basa sull'assunzione che le medie delle popolazioni siano uguali nei gruppi testati e determina se la variazione tra le medie campionarie dei gruppi sia maggiore di quella attesa per effetto del caso. Ciò indicherebbe che almeno una delle medie delle popolazioni è diversa dalle altre e che l'ipotesi nulla di uguaglianza delle medie deve essere rifiutata (Whitlock and Schluter 2010).

Questo test è stato utilizzato sia per verificare assenza di stratificazione del campione dovuta alla variabile quantitativa "età" sia per testare le associazioni genotipo/fenotipo.

Se non venisse preso in considerazione il problema della stratificazione potremmo rischiare di attribuire l'associazione di un fenotipo ad un dato gruppo genotipico che tuttavia potrebbe essere arricchito ad esempio di individui con età media molto diversa da quella degli altri gruppi. Pertanto non potremmo escludere con certezza l'ipotesi che tale associazione non sia influenzata dall'età.

Nei test di associazione genotipo/fenotipo l'ipotesi nulla è che le medie delle misurazioni fenotipiche rilevate per i campioni di alta quota non differiscono significativamente tra le categorie determinate dal genotipo. Pertanto se il test devia significativamente dall'ipotesi nulla si può assumere che il valore medio del fenotipo indagato è diverso in almeno uno dei gruppi, ossia in uno dei genotipi. Questo ci ha permesso di rilevare se ci

sono associazioni tra un determinato genotipo e il valore medio di una variabile fenotipica quantitativa.

Questi test sono stati effettuati utilizzando il software STATISTICA versione 7.1 (StatSoft).

5. IL CAMPIONE ANDINO NEL CONTESTO SUD-AMERICANO

Prima di procedere con lo studio dei geni candidati per l'adattamento all'alta quota, le popolazioni oggetto di questo studio sono state tipizzate ad uno dei marcatori più comunemente utilizzato per descrivere la variabilità genetica delle popolazioni umane, il DNA mitocondriale. Questo ci ha permesso di ottenere dati di variabilità genetica ad un locus considerato neutrale, dunque informativo circa la storia demografica delle nostre popolazioni, da poter confrontare con i dati ottenuti analizzando la variabilità genetica dei loci codificanti del nostro studio. Inoltre, data l'ampia disponibilità di dati di DNA mitocondriale in Sud America, è stato possibile verificare la struttura genetica in relazione a quella geografica del subcontinente includendo i nostri campioni nell'analisi.

La variabilità genetica delle popolazioni sud americane è strettamente connessa ai fenomeni storici e demografici che riguardano il popolamento del continente americano.

5.1 Il popolamento dell'America

Circa 100 mila anni fa una popolazione umana con caratteristiche anatomiche moderne originatasi in Africa, si è espansa fuori dal continente colonizzando rapidamente tutto il pianeta (Out of Africa) e rimpiazzando le forme umane arcaiche. Seguendo il percorso delle migrazioni dell'uomo in questo processo di colonizzazione, il continente americano risulterebbe l'ultimo continente ad essere popolato (Fig.12).

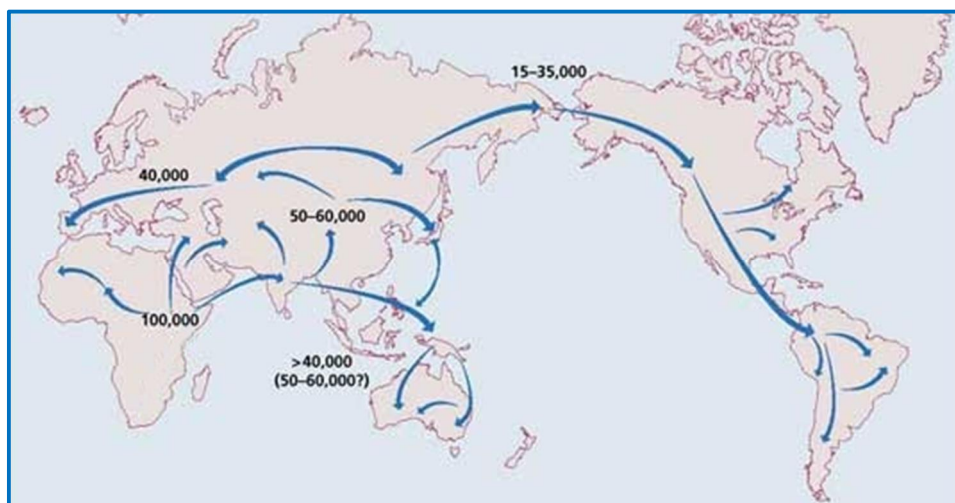


Fig.12 *Out of Africa* e colonizzazione del pianeta (Jobling et al. 2004b)

Secondo questo modello le popolazioni americane hanno avuto origine da popolazioni asiatiche che si spinsero verso la regione della Siberia in un periodo anteriore alla Last Glacial Maximum (LGM: 26,5-20 mila anni fa, Clark et al. (2009)).

A questo punto si apre un dibattito scientifico, ancora oggi oggetto di numerosi studi, circa le date di entrata nel continente americano, il numero di ondate migratorie, le modalità e i percorsi seguiti dalle popolazioni nella discesa verso l'estremo sud del continente (Rothhammer and Dillehay 2009).

Le prime ipotesi circa il modello di popolamento risalgono al 1932 quando fu scoperto in Nord America il sito archeologico Clovis datato circa 11 mila anni fa. Questa scoperta fece ipotizzare il cosiddetto modello "Clovis First" ossia un modello che prevede una singola ondata migratoria avvenuta al termine della LGM quando le terre erano ormai libere dai ghiacci.

Questo modello fu successivamente rivalutato sulla base di dati genetici relativi all'analisi del mtDNA e del cromosoma Y, e nuove scoperte archeologiche di cui la più nota è il sito di Monteverde, in Cile, risalente a 14500 anni fa, che anticipa l'entrata in America rispetto all'ipotesi di Clovis.

Il dibattito a questo punto si è concentrato sulla comprensione di come sia stato possibile per i primi popoli di cacciatori raggiungere e attraversare le regioni del Nord America in un periodo in cui metà del continente era ancora ricoperto dai ghiacci. Nel 1933 W.A. Johnston propose l'ipotesi dell'esistenza di un corridoio libero dai ghiacci definito appunto "*ice free corridor*", compreso tra i due principali ghiacciai chiamati Laurentide e Cordilleran. Tuttavia ancora oggi ci sono dubbi relativi alla sua reale esistenza (Rothhammer and Dillehay 2009).

Altri ricercatori ipotizzano che le prime migrazioni umane siano avvenute seguendo una rotta lungo la costa ovest del Nord America libera dai ghiacci già a partire da 16 mila anni fa (Wang et al. 2007). Ciò sarebbe in accordo con i reperti archeologici, come il sito di Monteverde, datati appunto intorno ai 14 mila anni fa (Dillehay et al. 2008).

Nonostante questa sia l'ipotesi più avvalorata molte domande restano aperte poiché una chiara comprensione della storia pre-coloniale e della genetica delle popolazioni native americane è resa particolarmente difficile dalla scarsità di reperti archeologici e di popolazioni native del continente. Ciò è principalmente dovuto alla colonizzazione del continente da parte degli Europei intorno al 1500 d.C, accompagnata dallo sterminio e dalla marginalizzazione di molti popoli nativi americani. Questo fenomeno pone il problema del mescolamento genetico avvenuto tra le popolazioni native americane e i popoli colonizzatori di origine europea.

5.2 Le popolazioni native del Sud America: aspetti genetici

Gli eventi demografici sopra descritti che hanno caratterizzato il continente americano hanno avuto un impatto rilevante nel modellamento della variabilità genetica di queste popolazioni. Dalla letteratura sappiamo che il livello di variabilità genetica delle popolazioni umane è in generale correlato alla distanza geografica di queste ultime dal continente africano dal quale hanno avuto origine. Numerosi studi hanno dimostrato che il livello di eterozigosità (utilizzato per stimare la variabilità genetica) si riduce progressivamente man mano che ci si allontana dall'Africa (Prugnolle et al. 2005, Ramachandran et al. 2005, Li et al. 2008) seguendo le rotte di colonizzazione del pianeta. Ciò significa che individui nativi del continente americano, essendo l'ultimo continente colonizzato, presentano di per se un basso livello di variabilità genetica rispetto ad individui provenienti da altri continenti e in particolare rispetto all'Africa. Inoltre, il periodo di isolamento in Beringia e la rapida espansione demografica durante le successive ondate migratorie per la colonizzazione dell'intero continente, hanno contribuito a determinare una progressiva riduzione della variabilità genetica interna secondo un gradiente nord-sud (Wang et al. 2007). Ciò è dimostrato anche dal numero ridotto di aplogruppi del mtDNA e del cromosoma Y riscontrati in queste popolazioni (paragrafo 5.3.1).

Focalizzando l'attenzione sulle popolazioni native del Sud America, oggetto di questo studio, non è possibile definire un pattern omogeneo di variabilità. Diversi studi basati sull'analisi del mtDNA, del cromosoma Y e di microsatelliti autosomici, hanno dimostrato

la presenza di differenze tra le popolazioni dell'est, prevalentemente situate nella regione delle foreste amazzoniche, e le popolazioni andine della regione ovest (Tarazona-Santos et al. 2001, Fuselli et al. 2003, Wang et al. 2007).

Nello specifico la variabilità intra-popolazione nella regione andina risulta elevata probabilmente come conseguenza del mantenimento di grandi dimensioni effettive delle popolazioni (N_e) e dell'elevato flusso genico tra loro. Una popolazione con una grande dimensione effettiva è meno soggetta all'effetto della deriva genetica, pertanto mantiene al suo interno un elevato livello di variabilità. Inoltre le frequenti migrazioni tra le popolazioni di questa regione determinano un continuo scambio (flusso genico) che porta ad omogenizzare il pool di variabilità (Fig.13).

Dal punto di vista socio culturale, l'elevato livello di similarità tra le popolazioni Andine può essere spiegato con un'agricoltura fiorente e alti livelli di struttura sociale ed economica che hanno favorito un aumento delle dimensioni delle popolazioni e dei contatti tra di esse (Stanish 2001, Tarazona-Santos et al. 2001). Questo è piuttosto

rilevante se si considera che generalmente l'ambiente di alta quota è considerato una barriera al flusso genico ed è associato a isolamento e differenziazione delle popolazioni.



Fig.13 Rappresentazione grafica dei pattern di variabilità genetica delle popolazioni dell'est e dell'ovest. La dimensione dei cerchi rappresenta la dimensione effettiva delle popolazioni mentre la dimensione delle frecce è proporzionale al livello di flusso genico. Tarazona-Santos et al. (2001)

Al contrario le popolazioni dell'Amazzonia e del Brasile mostrano piccole dimensioni effettive delle popolazioni e pertanto una riduzione della variabilità genetica per effetto della deriva. Inoltre il flusso genico tra le popolazioni della regione est è particolarmente ridotto determinando un'elevata variabilità tra popolazioni (Tarazona-Santos et al. 2001). Ciò può essere associato ad un progressivo isolamento dei piccoli villaggi che costituiscono questa regione a causa della fiorente vegetazione della foresta tropicale che ha determinato una barriera al flusso genico.

Per questi motivi le popolazioni andine sono ancora oggi oggetto di numerosi studi sia legati all'adattamento all'alta quota sia volti ad aggiungere tasselli importanti nella comprensione dei fenomeni demografici che hanno modellato la variabilità genetica di questo continente.

In questo studio per analizzare la variabilità genetica del campione andino e confrontarla con quella delle altre popolazioni native sud americane abbiamo tipizzato la regione ipervariabile I (HVRI) del DNA mitocondriale (mtDNA).

5.3 Il DNA mitocondriale

Il DNA mitocondriale umano (mtDNA) è una molecola circolare di circa 16,5 Kb composto da due eliche, una ricca in guanina (G-rich) e l'altra ricca in citosina (C-rich) che vengono definite rispettivamente catena pesante (H – heavy chain) e catena leggera (L – light chain) (Fig.14). L'mtDNA contiene 37 geni che codificano per 2 RNA ribosomiali (rRNA), 22 RNA di trasporto (tRNA) e 13 proteine che fanno parte dei complessi enzimatici deputati alla produzione di energia (fosforilazione ossidativa).

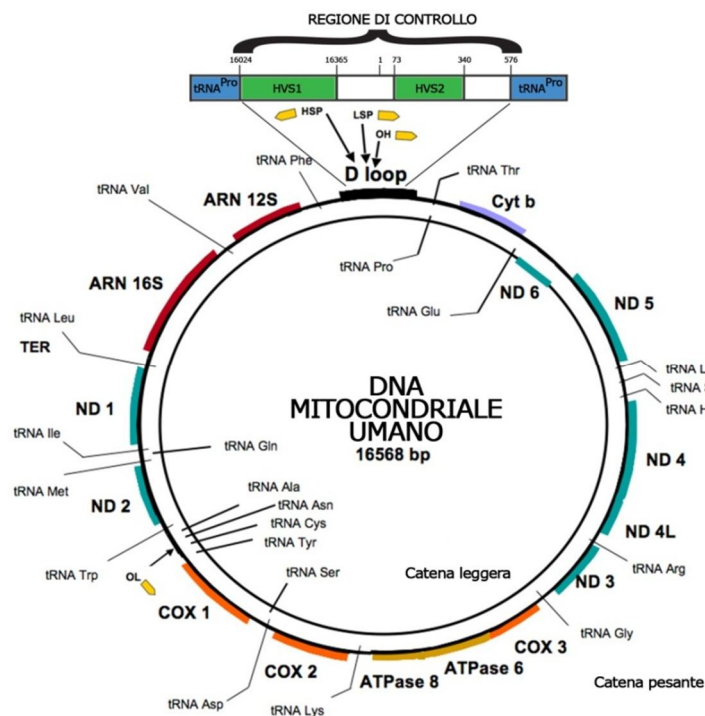


Fig.14 Rappresentazione grafica della molecola di DNA mitocondriale umano. Figura modificata da Bellance et al. (2009)

Il numero dei mitocondri varia dal tipo di cellula e, all'interno di ogni organello, si trovano in media da 2 a 10 copie di mtDNA, pertanto l'mtDNA è presente in un numero molto elevato di copie in ciascuna cellula. Questa caratteristica aumenta notevolmente l'efficienza nei processi di estrazione del DNA e permette di effettuare studi anche su reperti antichi (studi di DNA antico) o reperti degradati (genetica forense).

Inoltre le caratteristiche di "non ricombinazione", "ereditarietà solo materna" e "alto tasso di mutazione" ($3,4 \times 10^{-7}$) (Ingman et al. 2000) rendono questa molecola particolarmente adatta per gli studi di tipo evolutivo (Pakendorf and Stoneking 2005).

In particolare, regioni diverse del genoma mitocondriale presentano diversi livelli di variabilità a causa di varie pressioni selettive e strutture del DNA stesso.

Una regione molto importante è definita D-loop o regione di controllo. Il D-loop contiene al suo interno i due segmenti più polimorfici di tutto il genoma mitocondriale e sono chiamati regioni ipervariabili I e II (HVR I e HVR II). Queste due regioni vengono analizzate molto frequentemente in studi di genetica di popolazioni.

5.3.1 Distribuzione degli aplogruppi

Un modo per descrivere la variabilità genetica di un marcatore come l'mtDNA è quello di riassumerla in aplogruppi, dove per aplogruppo si intende un gruppo di sequenze simili, o aplotipi, anche se non necessariamente identiche, che condividono un antenato comune. Questo antenato e tutti gli aplotipi di un determinato aplogruppo sono caratterizzati da un particolare stato di uno o più determinati SNP.

La distribuzione dei diversi aplogruppi dell'mtDNA nelle diverse regioni geografiche del mondo è riassunta in figura 15.

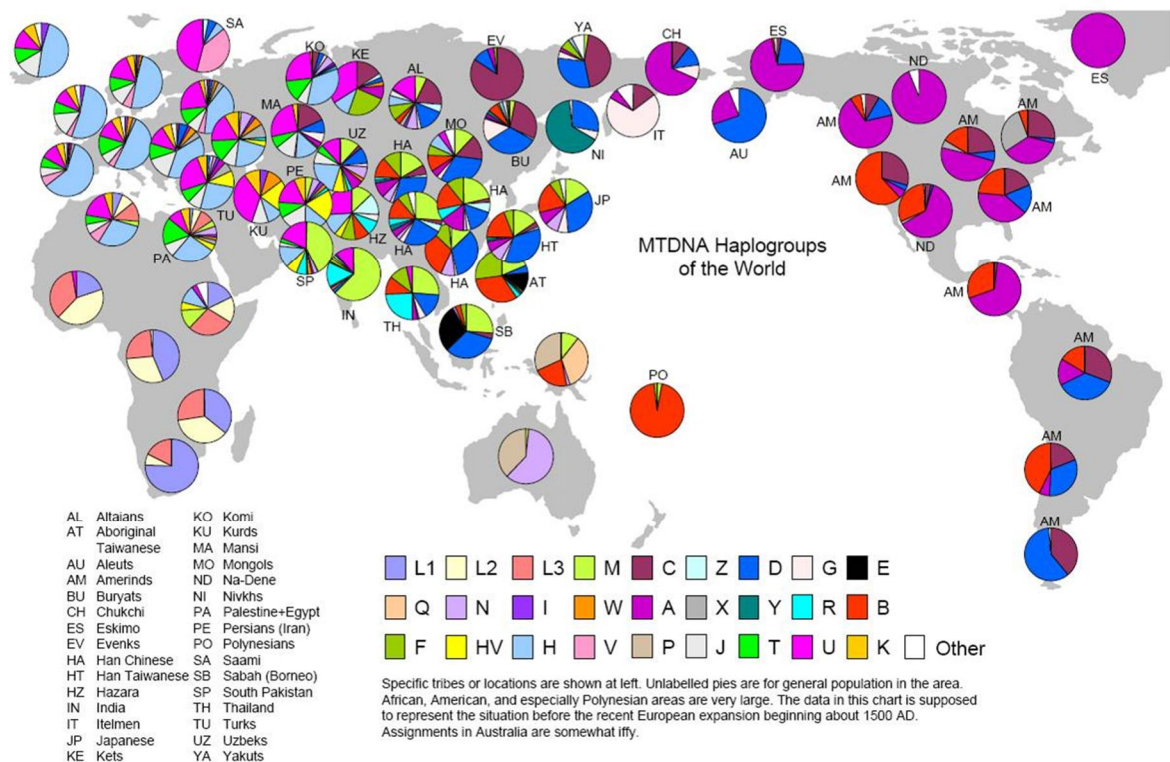


Fig.15 Rappresentazione grafica della distribuzione dei diversi aplogruppi dell'mtDNA nel mondo.
<http://www.scs.illinois.edu/~mcdonald/WorldHaplogroupsMaps.pdf>

Come si può notare, per il continente americano sono stati individuati quattro aplogruppi tipici, anche se non esclusivi: A, B, C, e D (Bonatto and Salzano 1997).

In alcuni lavori viene considerato anche l'aplogruppo X, anche se ha caratteristiche particolari: è presente a bassa frequenza solo nel Nord America, è assente in Siberia e ha un tempo di coalescenza più recente rispetto agli altri quattro aplogruppi (Fagundes et al. 2008).

Questi aplotipi rappresentano il 95-100% degli aplotipi mitocondriali presenti in America e sono considerati gli "aplotipi fondatori" in quanto sono stati rilevati anche in campioni antichi di individui Amerindi (Schurr and Sherry 2004).

Come descritto nei paragrafi 5.1 e 5.2 questa riduzione del numero di aplogruppi rispetto alle altre popolazioni del mondo si è probabilmente originata nella regione della Beringia, dove i primi insediamenti umani, già soggetti all'effetto fondatore, sono stati sottoposti ad una riduzione demografica in seguito alle avverse condizioni ambientali durante la Last Glacial Maximum, datata tra i 21.000 e i 18.000 anni fa (Fagundes et al. 2008).

Nel nostro studio abbiamo analizzato 360 bp della regione ipervariabile I del mtDNA. In particolare si tratta della sequenza che parte dalla base 16024 e termina alla base 16383 della sequenza di riferimento di Anderson (Anderson et al. 1981). Le analisi sono state effettuate su 209 individui dei 227 campionati; 18 individui sono stati esclusi dall'analisi a causa di relazioni di parentela con un membro dei soggetti investigati.

I siti polimorfici indagati per poter assegnare la sequenza di ciascun individuo all'aplogruppo di appartenenza sono riassunti in tabella 5.

Aplogruppo	16189 T/C	16217 T/C	16290 C/T	16319 G/A	16325 T/C	16327 C/T	16362 T/C
ref	T	T	C	G	T	C	T
A	-	-	T	A	-	-	-
B	C	C	-	-	-	-	-
C	-	-	-	-	C	T	-
D	-	-	-	A	C	-	C

Tab.5 Schema dei siti polimorfici dell'HVRI dell'mtDNA utilizzati per l'attribuzione degli aplogruppi (Achilli et al. 2008).

Per una assegnazione più precisa all'aplogruppo B in individui che presentavano solo la sostituzione 16189 T>C, abbiamo anche investigato la presenza di una delezione di 9 bp nella regione V del mtDNA tra i geni per la citocromo ossidasi II e il tRNA per la lisina (Cann and Wilson 1983). Tra i due geni sopra citati solitamente è presente una ripetizione in tandem della sequenza 5'-CCCCCTCTA-3' (Wrischnik et al. 1987) mentre la delezione di una copia di questa ripetizione è un marcatore chiave per l'attribuzione di un aplotipo all'aplogruppo B (Schurr et al. 1990). Tutti i campioni di dubbia appartenenza all'aplogruppo B hanno dato risultati positivi alla delezione di 9 bp e quindi sono stati confermati appartenenti al suddetto aplogruppo.

Tutti gli individui tipizzati in questo studio appartengono ad uno dei quattro principali aplogruppi tipici delle popolazioni native americane (Schurr and Sherry 2004) (Fig.16). Inoltre le frequenze di questi aplogruppi sono in accordo con quelle riscontrate per le popolazioni native americane in altri lavori (Schurr 2004) (Fig.17).

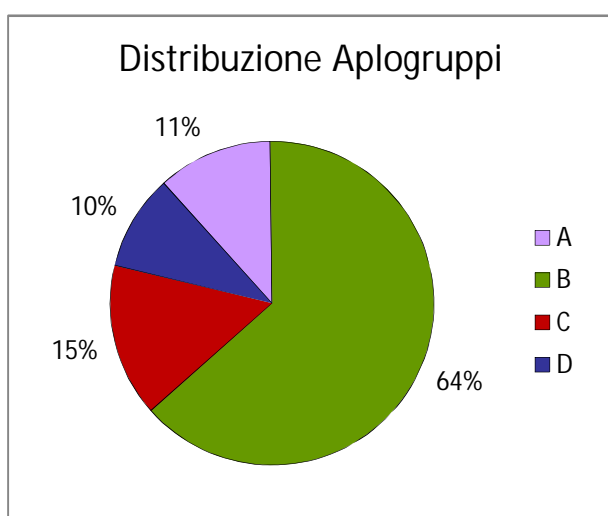


Fig.16 Distribuzione degli aplogruppi dell'mtDNA per tutti gli individui tipizzati in questo studio.

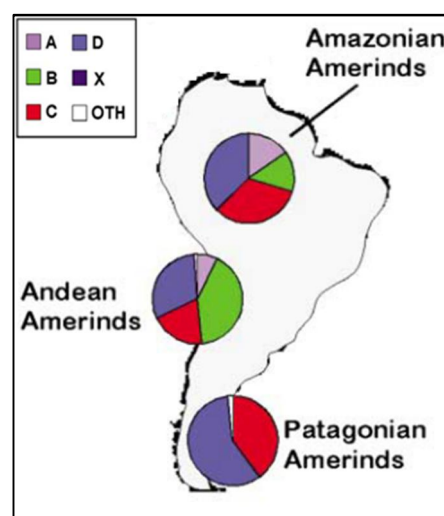


Fig.17 Distribuzione degli aplogruppi in Sud America (Schurr 2004)

5.3.2 Variabilità genetica intra-popolazione

Allo scopo di descrivere la variabilità genetica intra-popolazione sono stati calcolati alcuni indici di diversità genetica e sono state effettuate alcune stime del parametro mutazionale $\vartheta = 2nN_e\mu$ (Tab.6). Per una descrizione dettagliata vedere il paragrafo 4.2.3 di questa tesi.

mtDNA 16024-16383	Cusco	Cusibamba	Puno
N. di individui	60	18	131
N. di siti polimorfici	56	18	55
N. di aplotipi	43	9	60
Diversità aplotipica \pm SD	0,96 \pm 0,02	0,90 \pm 0,04	0,95 \pm 0,01
$\pi \pm$ SD ($\times 10^{-3}$)	17,98 \pm 9,57	18,61 \pm 10,30	17,10 \pm 9,06
$\theta_s \pm$ SD (per sites $\times 10^{-3}$)	33,36 \pm 9,80	14,54 \pm 5,95	28,04 \pm 7,39
Tajima's D	-1,84*	0,35	-1,50*
Fu's F_s	-25,40*	-0,002	-25,31*

Tab.6 Indici di variabilità intra-popolazione e test di neutralità calcolati per le popolazioni di questo studio.

*p-value < 0,05;

DS (Standard Deviation): deviazione standard

π : Diversità nucleotidica (Tajima 1983), basata sul numero medio di differenze tra coppie di sequenze sul totale dei siti.

θ_s (Watterson 1975), basato sul numero di siti polimorfici osservati (S).

In generale per le popolazioni andine osserviamo livelli abbastanza alti di diversità genetica interna.

Per alcune stime il campione di Cusibamba mostra valori di variabilità genetica più bassi rispetto agli altri campioni. Questo potrebbe essere spiegato dal fatto che si tratta di una popolazione isolata geograficamente e di piccole dimensione (180 unità familiari per un totale di non più di 1000 individui) e che per questi motivi ha subito maggiormente l'effetto della deriva genetica.

Ciò è evidente anche dai risultati dei test di neutralità D di Tajima e F_s di Fu che riportano valori negativi significativi e paragonabili per le popolazioni di Cusco e Puno, mentre per le popolazioni di Cusibamba tali valori non sono significativi e sono molto prossimi allo zero. Considerando esclusivamente una interpretazione di tipo demografico, dato che

ipotizziamo assenza di selezione nella regione HVR, i risultati di questi due test evidenziano una recente espansione delle popolazioni di Cusco e Puno, mentre tale segnale è assente per la popolazione di Cusibamba.

La variabilità genetica di queste popolazioni è stata successivamente confrontata con quella di altre popolazioni sud americane tipizzate per la stessa regione HVRI dell'mtDNA da Fuselli et al. (2003) e aggiornate dal gruppo del professor Eduardo Tarazona Santos dell'Università di Belo Horizonte, in Brasile (Tesi master di Luciana Werneck Zuccherato) (Fig.18).

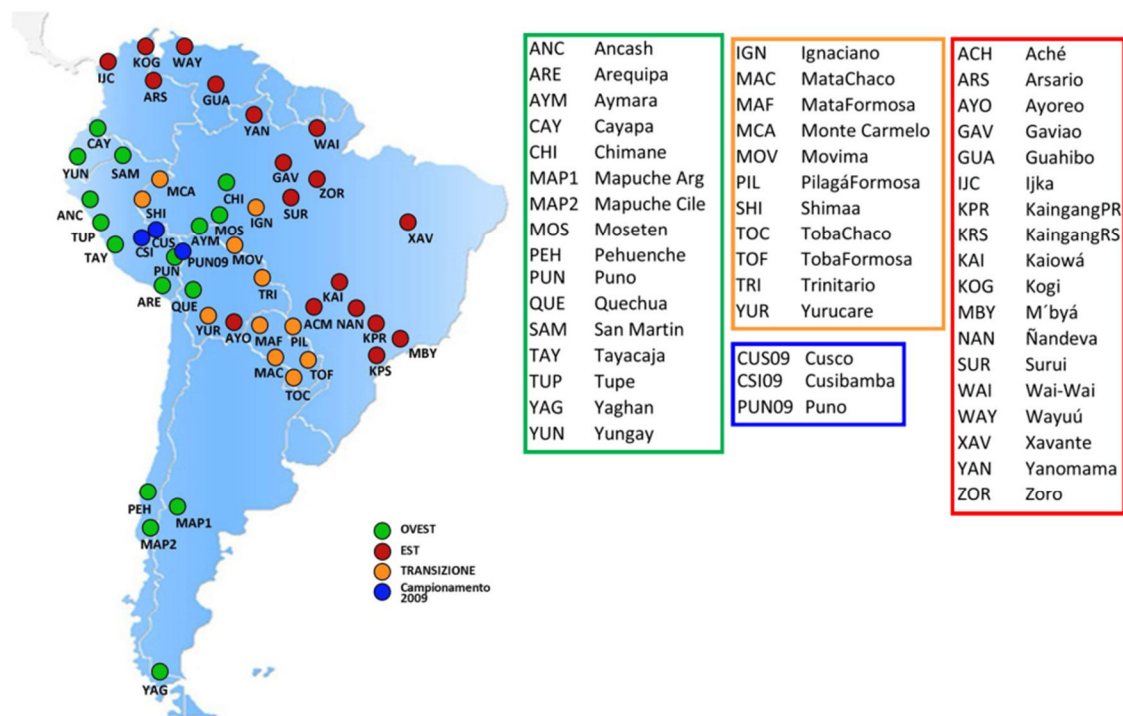


Fig.18 Localizzazione geografica delle popolazioni sudamericane utilizzate per le analisi relative all'mtDNA. A destra sono riportati i nomi delle popolazioni indicate nella cartina geografica.

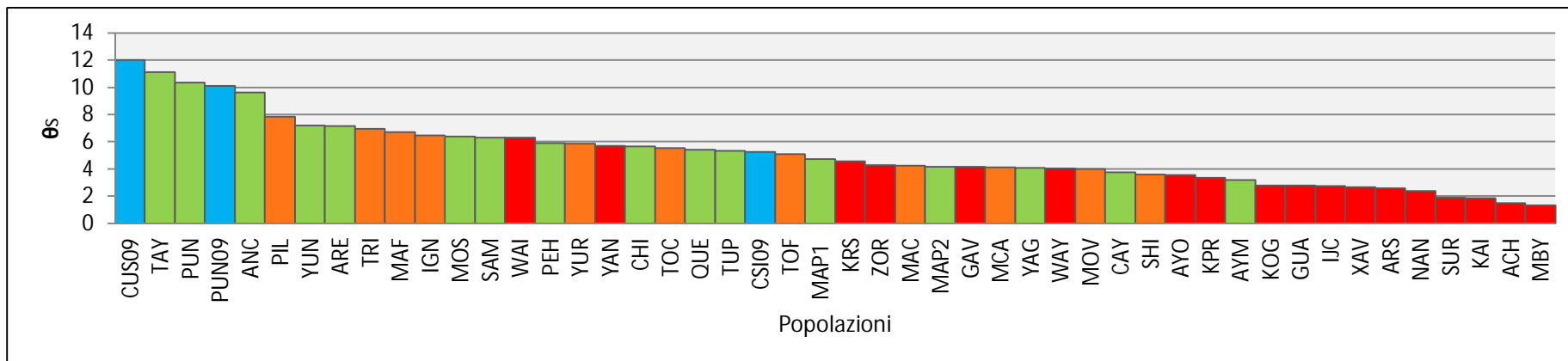
Come si vede in figura 18 la posizione geografica delle popolazioni consente una buona copertura del territorio. Le popolazioni possono essere suddivise secondo un criterio geografico in popolazioni dell'ovest o andine, popolazioni dell'est, e popolazioni collocate in una zona detta di transizione intermedia tra la catena montuosa delle Ande e la foresta amazzonica.

Per tutte le popolazioni oggetto di studio sono stati calcolati gli indici di diversità e i test di neutralità già stimati per le popolazioni da noi campionate.

I nostri risultati mostrano come in generale le popolazioni Andine, tra cui anche le popolazioni di Cusco e Puno, confermino gli andamenti di variabilità genetica osservati in studi precedenti.

Confrontando la variabilità genetica interna delle popolazioni di questo dataset è possibile osservare come le popolazioni dell'est mostrino livelli di variabilità più bassi rispetto alle popolazioni della regione andina (Fig.19).

Le popolazioni cosiddette di transizione mostrano livelli di variabilità genetica intermedia, sebbene più simile a quella delle popolazioni dell'ovest, da cui si potrebbe dedurre un mantenimento di dimensioni effettive non molto ridotte come quelle delle popolazioni amazzoniche. Tuttavia non si conosce molto sulla loro storia e sulle loro origini. Attualmente il gruppo del professor Tarazona-Santos dell'Università di Belo Horizonte in Brasile ha avviato degli studi più approfonditi per comprendere la storia demografica di queste popolazioni, anche utilizzando marcatori nucleari.

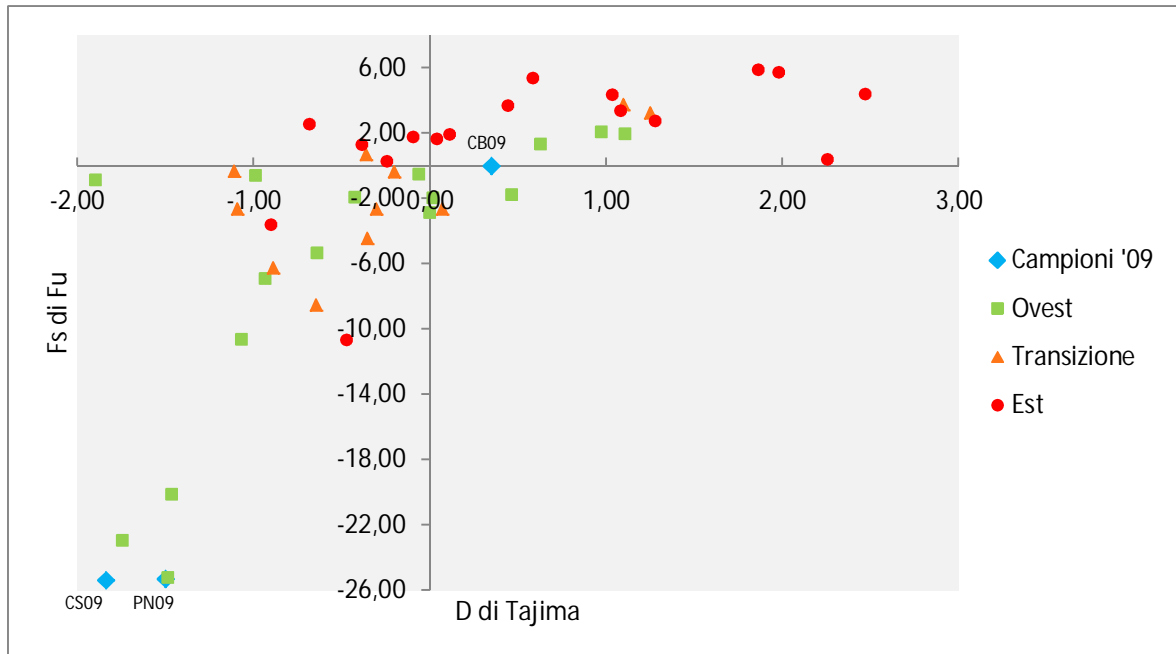


ANC	Ancash
ARE	Arequipa
AYM	Aymara
CAY	Cayapa
CHI	Chimane
MAP1	Mapuche Arg
MAP2	Mapuche Cile
MOS	Moseten
PEH	Pehuenche
PUN	Puno
QUE	Quechua
SAM	San Martin
TAY	Tayacaja
TUP	Tupe
YAG	Yaghan
YUN	Yungay
IGN	Ignaciano
MAC	MataChaco
MAF	MataFormosa
MCA	Monte Carmelo
MOV	Movima
PIL	PilagáFormosa
SHI	Shimaa
TOC	TobaChaco
TOF	TobaFormosa
TRI	Trinitario
YUR	Yurucare
CUS09	Cusco
CSI09	Cusibamba
PUN09	Puno
ACH	Aché
ARS	Arsario
AYO	Ayoreo
GAV	Gaviao
GUA	Guahibo
IJC	Ijka
KPR	KaingangPR
KRS	KaingangRS
KAI	Kaiowá
KOG	Kogi
MBY	M'byá
NAN	Ñandeva
SUR	Surui
WAI	Wai-Wai
WAY	Wayuú
XAV	Xavante
YAN	Yanomama
ZOR	Zoro

Fig.19 Rappresentazione grafica delle popolazioni ordinate per valori crescenti di θ_s .

Sotto la figura sono riportati i nomi delle popolazioni con le relative abbreviazioni. Il colore verde indica popolazioni dell'ovest, il colore arancione le popolazioni di transizione, il colore rosso le popolazioni dell'est e in azzurro sono indicate le popolazioni dell'ovest campionate in questo studio.

In figura 20 sono rappresentati i valori dei test di neutralità D di Tajima (Tajima 1989) e F_s di Fu (Fu and Li 1993) descritti in dettaglio nel paragrafo 4.2.3.2.



5.3.3 Variabilità genetica tra popolazioni

La distanza genetica tra coppie di popolazioni è stata stimata mediante il calcolo di una matrice di F_{ST} che è stata poi visualizzata graficamente attraverso un multidimensional scaling (MDS) (Fig.21).

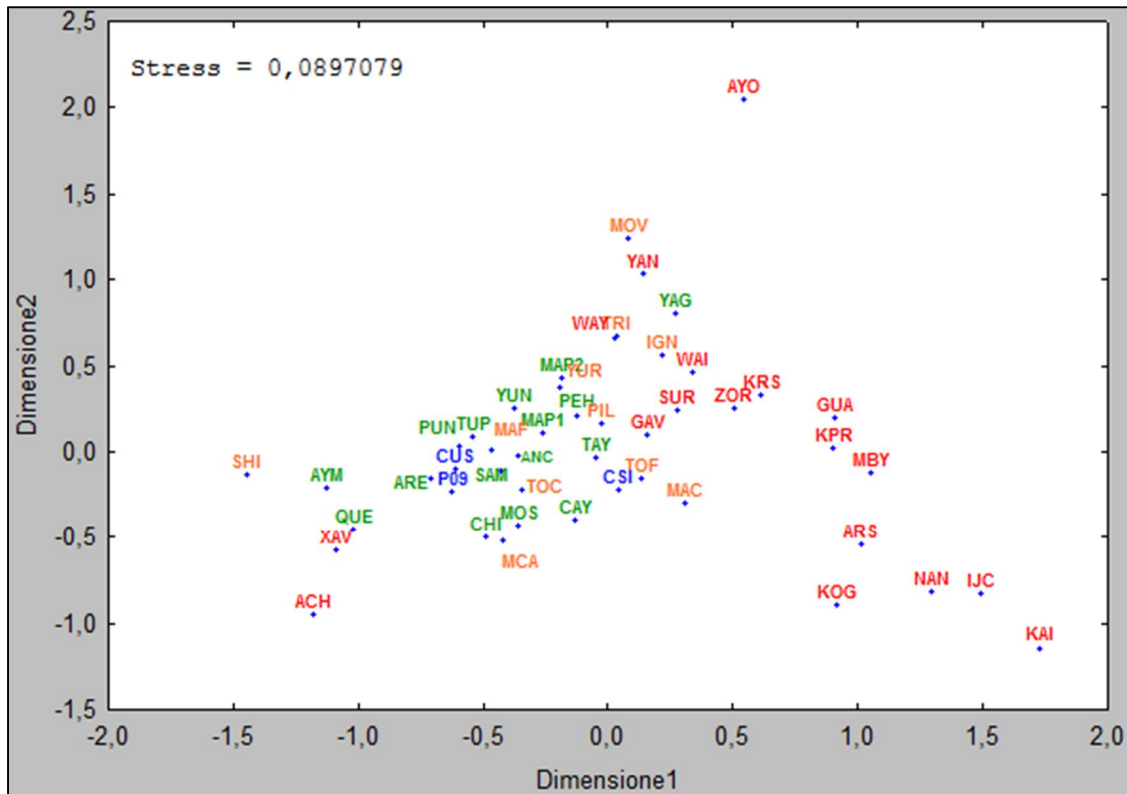


Fig.21 Multidimensional scaling basato sulla matrice di distanze genetiche F_{ST} . In rosso sono indicate le popolazioni della regione est, in verde quelle dell'ovest e in arancione le popolazioni della zona di transizione. Le popolazioni campionate da noi sono indicate in blu.

Oltre a mettere in luce da un'altra prospettiva la struttura est-ovest, l'MDS mostra come le popolazioni della regione orientale (in rosso) siano meno omogenee e molto distribuite lungo entrambi gli assi. Viceversa, si nota una omogeneità delle popolazioni occidentali, che include anche la maggior parte delle popolazioni di transizione, di cui fanno parte anche i nostri tre campioni andini.

6. IDENTIFICAZIONE ED ANALISI DEI GENI CANDIDATI ALLO STUDIO DELL'ADATTAMENTO GENETICO UMANO ALL'ALTA QUOTA

Dopo aver descritto alcune delle caratteristiche demografiche delle popolazioni andine oggetto di questo studio grazie all'analisi del DNA mitocondriale, siamo passati allo studio delle regioni del genoma nucleare per analizzare la variabilità dei geni candidati alla risposta alle basse pressioni di ossigeno e rilevare eventuali segnali dell'azione della selezione naturale a questi loci.

6.1 Identificazione dei geni candidati

Lo studio di fenotipi o patologie complesse mediante l'approccio dei geni candidati risulta efficace se la selezione a priori dei geni e delle regione genomiche da analizzare viene effettuata attraverso metodi accurati.

Il metodo classico è rappresentato dalla ricerca in letteratura, tuttavia tale metodo comporta problematiche legate all'accuratezza della selezione, data la vastità del database "letteratura" e la necessità di trovare indicazioni e suggerimenti per indirizzare il proprio lavoro verso una strada nuova, originale e non già percorsa.

Recentemente sono stati sviluppati metodi bioinformatici in grado di integrare le informazioni provenienti dalla letteratura con quelle provenienti da più banche dati biologiche (Rzhetsky et al. 2004, Kent et al. 2005, Aerts et al. 2006, Jensen et al. 2009) per creare una scala di priorità basata su dei p-value che definiscono l'importanza di un determinato gene in relazione alla patologia o al fenotipo oggetto di studio. Questi metodi vengono appunto definiti "prioritization methods".

Uno dei software bioinformatici che integra le informazioni provenienti dalla maggior parte dei database biologici attualmente disponibili è rappresentato da ENDEAVOUR (Aerts et al. 2006, Tranchevent et al. 2008).

6.1.1 Endeavour

Endeavour (<http://homes.esat.kuleuven.be/~bioiuser/endeavour>) è un software che sfrutta i "prioritization methods" basandosi principalmente sulla stima dei livelli di similarità tra un potenziale gene candidato e alcuni geni già noti per essere coinvolti nel fenotipo o nella patologia oggetto di studio (Fig.22).

Questi geni, definiti "*training genes*", devono essere forniti al programma prima di procedere alla fasi successive. Per identificarli si possono sfruttare le informazioni presenti in letteratura inserendo direttamente nel programma l'identificativo dei geni o la regione cromosomica.

In alternativa si possono sfruttare le informazioni provenienti dai database:

- Gene Ontology (<http://www.geneontology.org>) ,
- KEGG (<http://www.genome.jp/kegg/>)
- OMIM (<http://www.ncbi.nlm.nih.gov/omim>) .

Nel caso di Gene Ontology o KEGG saranno inseriti tutti i geni annotati in banca dati in relazione ad una parola chiave fornita dall'utente (ad esempio "hypoxia") mentre nel caso di OMIM tutti i geni che contengono quella parola chiave, nella loro descrizione.

A questo punto è necessario selezionare le banche dati che il programma dovrà utilizzare per estrapolare le caratteristiche riguardanti i geni noti. Queste banche dati vengono definite *sottomodelli* (pezzi del puzzle) e andranno a definire un *modello* generale (il puzzle) costituito dall'insieme di tutti i database. Per ogni sottomodello i geni noti si arricchiscono di molte informazioni che andranno tutte insieme a comporre il file definito "*Training genes*".

Le banche dati utilizzate da Endeavour sono le seguenti:

- letteratura (EntrezGene, <http://www.ncbi.nlm.nih.gov/gene>),
- informazioni funzionali (Gene Ontology, <http://www.geneontology.org>),
- dati di espressione (Atlas gene expression, <http://www.ebi.ac.uk/gxa/>),
- dati EST (Ensembl, <http://www.ensembl.org>),
- domini proteici (InterPro, <http://www.ebi.ac.uk/interpro/>),
- interazioni interproteiche (Biomolecular Interaction Network Database – BIND, <http://bond.unleashedinformatics.com>),

- pathways (Kyoto encyclopedia of Genes and Genomes – KEGG, <http://www.genome.jp/kegg>),
- domini cis-regolatori (TOUCAN),
- motivi trascrizionali (TRANSFAC, <http://www.gene-81regulation.com>)
- similarità di sequenza (BLAST, <http://blast.ncbi.nlm.nih.gov/Blast.cgi>)

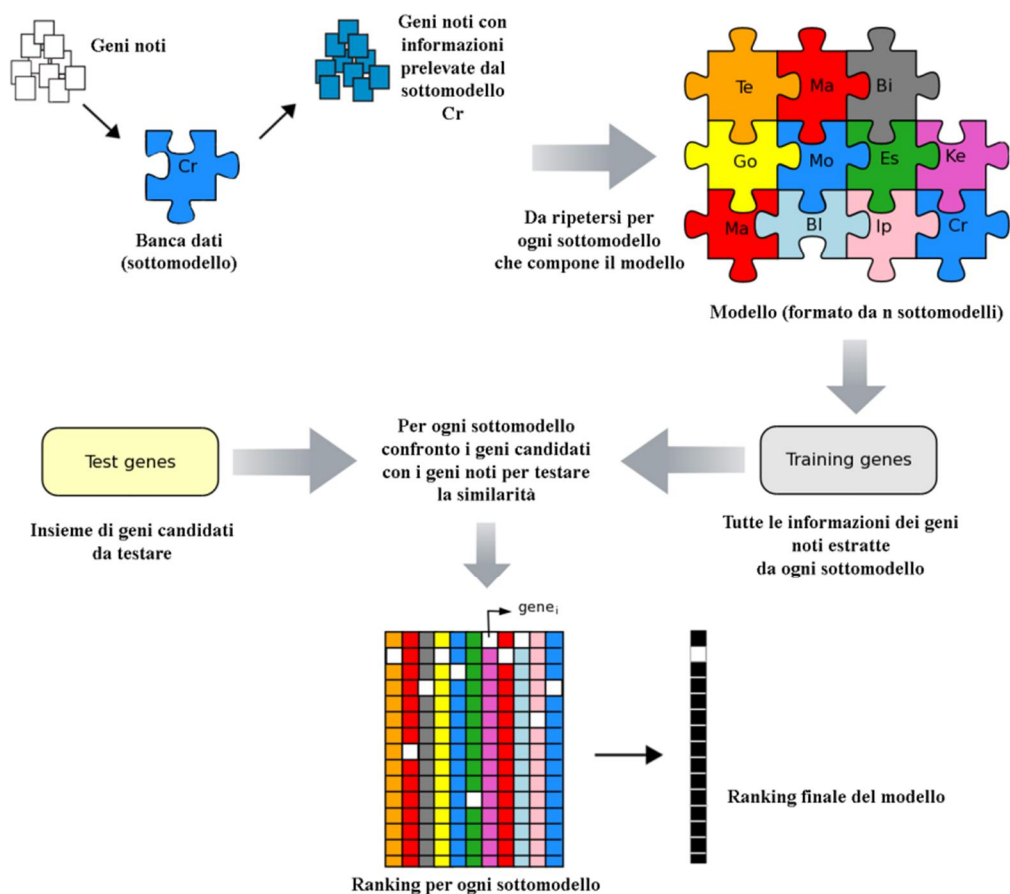


Fig.22 Funzionamento di *Endeavour*.

Separatamente si crea il secondo file di input, "Test genes", in cui si inseriscono i geni candidati da testare. Il programma, per ogni sottomodello, confronta i geni noti con i geni candidati per testare la similarità. In output viene fornito un ranking per ogni sottomodello in cui si può osservare la posizione dei geni candidati rispetto ai geni noti. Successivamente tutti i diversi rankings vengono fusi insieme per creare una classifica

unica del modello completo (Aerts et al. 2006, Tranchevent et al. 2008). Accanto alla visualizzazione grafica, è presente anche una tabella che associa ad ogni gene un p-value sia per ogni sottomodello analizzato sia per il modello completo.

I geni candidati selezionati tramite lo studio della letteratura (Hochachka and Rupert 2003, Hagg and Wennstrom 2005, Appenzeller et al. 2006, Brown et al. 2006, Rajput et al. 2006, Semenza 2007a, Breen et al. 2008, Leon-Velarde and Mejia 2008, Stobdan et al. 2008, Strohl 2008) sono stati testati con questo approccio per verificare la similarità con gruppi di geni noti tipici dei processi ipossici.

I risultati hanno confermato le aspettative collocando i geni di nostro interesse sempre nei primi posti della classifica.

Tali geni, elencati in tabella 7, fanno parte in modo diretto o indiretto della via metabolica di HIF descritta nel paragrafo 1.4.

Locus	Cromosoma	Funzione della proteina
<i>HIF1A</i>	14q21-q24	Fattori di trascrizione. Regolano la trascrizione di geni coinvolti nella risposta all'ipossia.
<i>HIF2A</i>	2p21-p16	
<i>EPO</i>	7q22	Target di HIF. Promuove la differenziazione degli eritrociti e la sintesi di emoglobina
<i>VHL</i>	3p26-p25	Coinvolto nella degradazione di HIF.
<i>NOS3</i>	7q36	Responsabile della sintesi di ossido nitrico (vasodilatatore) insieme a iNOS (NOS2A) e nNOS.
<i>VEGFA</i>	6p12	Target di HIF. Fattore di crescita coinvolto nella angiogenesi e nella crescita delle cellule endoteliali.
<i>EDN1</i>	6p24.1	Target di HIF. Codifica per una proteina che una volta processata proteoliticamente dà origine a dei peptide con funzione di vasocostrittori.
<i>EGLN1</i>	1q42.1	Codificano per le prolil idrossilasi che favoriscono la degradazione di HIF.
<i>EGLN2</i>	19q13.2	
<i>EGLN3</i>	14q13	

Tab.7 Geni candidati

7. EVOLUZIONE MOLECOLARE DEI GENI CANDIDATI

"[...] But if variations useful to any organic being do occur, assuredly individuals thus characterized will have the best chance of being preserved in the struggle for life; and from the strong principle of inheritance they will tend to produce offspring similarly characterized. This principle of preservation, I have called, for the sake of brevity, Natural Selection."

Charles Darwin, L'origine delle specie (1859), capitolo 4.

La principale opera di Charles Darwin in cui viene descritta per la prima volta la teoria dell'evoluzione degli organismi viventi nasce dall'osservazione e dal confronto di diverse specie in relazione a fattori ambientali. Questo metodo comparativo rappresenta ancora oggi il metodo più utilizzato per studiare i processi di adattamento per selezione naturale.

Alla base dei processi evolutivi ci sono cambiamenti che avvengono a livello della sequenza del DNA. Tali cambiamenti possono portare all'acquisizione di nuove caratteristiche morfologiche o fisiologiche che potranno diffondersi nella popolazione per effetto della selezione naturale e/o della deriva genetica. Ciò che maggiormente influisce nel determinare una o l'altra cosa è l'ambiente.

Pertanto confrontare le sequenze di DNA di specie o popolazioni che si sono evolute in ambienti che differiscono per un determinato fattore (ad esempio specie che vivono in ambienti a basse pressioni di ossigeno e specie che vivono in ambienti a normali pressioni di ossigeno) permette di identificare la presenza di differenze genetiche che possono aver influito sull'adattamento di quella specie (o popolazione) alle specifiche condizioni ambientali in cui si trova. A maggior ragione ciò sarà possibile se le sequenze di DNA che si confrontano appartengono a geni noti per svolgere un ruolo importante nel processo di adattamento al fattore ambientale in studio.

In questa tesi il metodo comparativo è stato applicato allo studio dei geni candidati all'adattamento alle basse pressioni di ossigeno. Tuttavia prima di procedere al confronto delle sequenze dei geni candidati a livello popolazionistico abbiamo effettuato uno studio a livello interspecifico. Questa analisi preliminare ha permesso di capire sotto quali

pressioni selettive si sono evoluti questi geni e se esistono differenze a livello interspecifico imputabili all'adattamento ad ambienti che differiscono per pressione di ossigeno nell'aria. Tutto ciò allo scopo di indirizzare le analisi successive, a livello intrapopolazionistico, verso specifiche regioni dei geni candidati.

7.1 La teoria neutrale e la stima dei tassi di sostituzione

"Calculating the rate of evolution in terms of nucleotide substitutions seems to give a value so high that many of the mutations involved must be neutral ones" (Kimura 1968).

Secondo la teoria neutrale dell'evoluzione sviluppata da Motoo Kimura nel 1968 la maggior parte della variabilità a livello molecolare è selettivamente neutrale e viene mantenuta nelle popolazioni grazie ad un equilibrio tra mutazione e deriva genetica.

Soltanto la variabilità genetica che determina un effetto sulla fitness, ossia sulla capacità di un organismo di sopravvivere e riprodursi, sarà soggetta all'azione della selezione naturale. Nel caso in cui le mutazioni producano un effetto svantaggioso sulla fitness saranno mantenute a basse frequenze per effetto della selezione negativa, o purificante. Al contrario nel caso in cui le mutazioni portino un vantaggio selettivo e favoriscano un migliore adattamento della specie all'ambiente, saranno mantenute a frequenza elevata (in alcuni casi portate a fissazione) per effetto della selezione positiva, adattativa (Kimura 1983).

La teoria neutrale ha ricevuto numerose conferme sperimentali ed oggi è utilizzata come "ipotesi nulla" nei test, definiti appunto *test di neutralità*, utilizzati negli studi di evoluzione molecolare. Questi test applicati ai dati reali sono sufficientemente potenti da rifiutare l'ipotesi nulla di neutralità permettendo di rilevare l'azione della selezione naturale in determinati geni. In particolare alcuni test utilizzati in studi interspecifici permettono di distinguere in modo inequivocabile l'azione della selezione positiva (adattativa).

I casi più convincenti di evoluzione molecolare adattativa sono stati identificati attraverso dei test che confrontano i tassi di sostituzione sinonime (d_S) e non-sinonime (d_N) nelle

sequenze codificanti del DNA (Yang and Bielawski 2000), dove per tasso di sostituzione si intende il numero di cambiamenti che raggiungono la fissazione nell'unità di tempo.

Considerando una sequenza di DNA codificante, per *sostituzioni sinonime* si intendono tutte quelle sostituzioni nucleotidiche che non hanno effetto sulla proteina codificata, non causando cambiamenti amminoacidici. Per questo vengono anche definite silenti. Le *sostituzioni non-sinonime* invece causano un cambiamento dell'amminoacido codificato dal codone in cui avvengono e pertanto sono in grado di produrre cambiamenti a livello proteico che possono essere selezionati positivamente o negativamente.

Poiché le sostituzioni sinonime non dovrebbero essere sottoposte a selezione naturale, il tasso di sostituzioni sinonime può essere considerato il tasso di sostituzione neutrale. Per questo motivo si osserva un simile tasso di sostituzione sinonime in diversi geni, mentre il tasso di sostituzioni non-sinonime è gene specifico. Il rapporto fra queste due misure può quindi essere utilizzato per valutare il tipo di selezione che ha agito sul gene di interesse e, di conseguenza, la sua evoluzione molecolare.

Per stimare questi tassi di sostituzione bisogna considerare tre principali problemi:

1. in un codone il numero di siti nucleotidici che mutati producono un cambiamento sinonimo è inferiore a quello dei siti che mutati producono un cambiamento non-sinonimo. Questa caratteristica si estende dunque a tutta la sequenza codificante e, per questo motivo, il tasso di sostituzione sinonimo deve essere definito come il numero di sostituzioni sinonime per sito sinonimo (d_S), e lo stesso varrà per i siti non-sinonimi (d_N);
2. definire il numero di siti sinonimi e non in un codone è complicato dal fatto che mutazioni diverse allo stesso sito possono dare origine ad un cambiamento sinonimo o ad uno non sinonimo:
GAT (Asp) > GAC (Asp) cambiamento in terza base sinonimo
GAT (Asp) > GAA (Glu) cambiamento in terza base non-sinonimo;

3. se due codoni differiscono per più di un cambiamento, si possono ipotizzare diversi percorsi attraverso i quali si è giunti alla situazione presente. Ad esempio, se confrontando due sequenze notiamo che la prima possiede il codone TTT alla stessa posizione in cui la seconda presenta il codone GTA, possiamo presumere almeno due percorsi che implicano un diverso numero di cambiamenti sinonimi e non-sinonimi:

TTT (Phe) > GTT (Val) > GTA (Val) = 1 non-sinonimo + 1 sinonimo

TTT (Phe) > TTA (Leu) > GTA (Val) = 2 non-sinonimi

Esistono diversi modi per stimare d_S e d_N in modo da considerare i problemi elencati sopra. Uno dei metodi più utilizzati in letteratura, il metodo di Nei e Godjobori (1986) calcola il numero di sostituzioni sinonime e non-sinonime sui siti *potenzialmente* sinonimi e non-sinonimi assumendo uguale probabilità per qualunque cambiamento nucleotidico (Nei and Godjobori 1986).

Per chiarire cosa si intende per siti potenzialmente sinonimi e non, facciamo un esempio. Consideriamo il codone codificante per la fenilalanina TTT: in questo caso, tutti i cambiamenti in prima e seconda posizione risultano in una mutazione non-sinonima, mentre i cambiamenti in terza base risultano in due mutazioni non-sinonime (TTA e TTG) ed una sinonima (TTC). Per cui si otterrà:

$$S = 0 + 0 + (1/3) = 1/3 \quad \text{mentre} \quad N = 3 - s = 8/3$$

Questo metodo è stato successivamente modificato per includere correzioni che considerano, ad esempio, che le transizioni (cambiamenti purina-purina o pirimidina-pirimidina) sono più comuni delle trasversioni (cambiamenti purina-pirimidina) ((Nei and Kumar 2000) capitolo 4 pag.57).

In alternativa, Goldman e Yang (1994) hanno sviluppato un metodo basato sulla massima verosimiglianza che è stato utilizzato in questa tesi e che verrà descritto più in dettaglio nel prossimo paragrafo (Goldman and Yang 1994).

Una volta ottenuta la stima di:

- d_S = numero di sostituzioni sinonime osservate sul totale dei siti sinonimi
- d_N = numero di sostituzioni non-sinonime osservate sul totale dei siti non-sinonimi

sarà possibile calcolare il rapporto tra queste due grandezze (d_N/d_S) definito omega (ω):

$$\omega = d_N/d_S$$

Questo indice viene calcolato per studiare l'evoluzione dei geni in analisi comparative.

In condizioni di neutralità il tasso di sostituzione sinonimo sarà uguale al tasso di sostituzione non-sinonimo e pertanto il valore di ω sarà uguale ad uno

$$d_N = d_S \rightarrow \omega = d_N/d_S = 1 \rightarrow \text{NEUTRALITÀ}$$

In caso di selezione negativa, purificante, il rapporto tenderà ad assumere valori inferiori ad uno perché un cambiamento non-sinonimo deleterio sarà fissato ad un tasso minore rispetto ad un cambiamento sinonimo.

$$d_N < d_S \rightarrow \omega = d_N/d_S < 1 \rightarrow \text{SELEZIONE PURIFICANTE (o NEGATIVA)}$$

In caso di selezione positiva, diversificante il rapporto sarà maggiore di uno perché il tasso di sostituzioni che determinano cambiamenti amminoacidici sarà maggiore di quello delle sostituzioni silenti

$$d_N > d_S \rightarrow \omega = d_N/d_S > 1 \rightarrow \text{SELEZIONE ADATTATIVA, DIVERSIFICANTE (o POSITIVA)}$$

((Nei and Kumar 2000) pag51-71).

Quindi stimare il valore di omega permette di definire quale delle possibili ipotesi circa l'evoluzione di una determinata sequenza codificante sia la più probabile.

7.2 PAML: Phylogenetic Analysis by Maximum Likelihood

Uno dei metodi più potenti utilizzati per la stima di ω sfrutta l'approccio di massima verosimiglianza (Maximum-Likelihood) (Goldman and Yang 1994). In generale i metodi di massima verosimiglianza si basano sulla *funzione di likelihood* secondo la quale è possibile stimare i valori dei parametri (ignoti) che più verosimilmente hanno prodotto i dati che stiamo osservando. Il valore del parametro che massimizza la verosimiglianza con i dati è definito come stima di massima verosimiglianza (Maximum Likelihood Estimate, MLE).

Questo metodo è implementato nel programma PAML (versione 4.4a) (Phylogenetic Analysis by Maximum Likelihood) (Yang 1997, Yang 2007b).

Basandosi su un modello specifico di sostituzione di nucleotidi per i 61 codoni del codice genetico universale (i codoni di stop vengono esclusi), il metodo è in grado di stimare dai dati, ossia dalla coppia di sequenze codificanti che stiamo analizzando, alcuni parametri come il tempo di divergenza tra le sequenze, il tasso di transizione/trasversione, e il valore di ω .

In particolare il modello assume che il cambiamento istantaneo, riferito cioè al singolo passaggio, di un codone q_i col codone q_j venga descritto dalle seguenti equazioni:

$q_{ij} = 0$ se il passaggio da i a j comporta più di un cambiamento nucleotidico (esempio AAG > ACT, 2 cambiamenti nello stesso istante, molto improbabile)

$q_{ij} = \pi_j$ per le trasversioni sinonime, dove π_j è la frequenza relativa del codone j -esimo

$q_{ij} = k\pi_j$ per le transizioni sinonime, dove k è il rapporto tra le transizioni e le trasversioni

$q_{ij} = \omega \pi_j$ per le trasversioni non-sinonime, dove ω è d_N/d_S

$q_{ij} = \omega k \pi_j$ per le transizioni non-sinonime

Ci sarebbero 61 parametri per π_j , ma possiamo stimare la frequenza di ogni codone da quella osservata se assumiamo che le frequenze dei codoni siano in equilibrio e se il numero di codoni utilizzato è grande. Quindi restano da stimare k ed ω con la massima verosimiglianza.

Il calcolo della massima verosimiglianza per sequenze multiple può essere considerato come un'estensione del calcolo per una coppia di sequenze (Yang 2007a).

PAML si serve di diversi modelli che permettono di stimare i valori di omega secondo diverse ipotesi. I modelli più semplici stimano omega medio per l'intero gene e non sono considerati dei test per individuare la selezione positiva, ma solo dei metodi per verificare se il gene nella sua interezza, in quel determinato gruppo di specie, si sia evoluto in condizioni di neutralità o sotto pressioni selettive differenti.

In una proteina che svolge un ruolo funzionale molti amminoacidi sono sottoposti a vincoli strutturali e funzionali che ne impediscono la variabilità e pochi di essi potranno subire sostituzioni tali da essere selezionate positivamente (Yang et al. 2005). Quindi stimare un unico valore medio di omega per l'intera sequenza non permetterebbe di rilevare gli eventuali siti sottoposti all'azione della selezione positiva.

La potenza di PAML sta nella sua capacità di implementare modelli più complessi che permettono di identificare anche singoli siti amminoacidici sottoposti a selezione in un contesto generalmente molto conservato come atteso secondo la teoria neutrale (*site models*).

PAML contiene modelli che permettono anche di testare l'ipotesi che alcune linee filogenetiche, scelte sulla base di un'ipotesi effettuata a priori, si siano evolute in modo differente da altre per il gene analizzato (*branch models*).

Inoltre diversamente da altri metodi che per stimare d_N e d_S nella filogenesi utilizzano una sequenza ancestrale ricostruita con metodi di parsimonia come se fosse un dato osservato, PAML sfrutta l'approccio di massima verosimiglianza per stimare tra tutte le possibili sequenze ancestrali ad ogni nodo interno dell'albero quella che massimizza il valore di verosimiglianza (Yang 1998).

Generalmente i modelli applicati dal programma si suddividono in un modello nullo e un modello alternativo, che contiene le stesse assunzioni del modello nullo, ma ipotizza in più la possibilità di selezione positiva nel caso dei *site models*, o di diverso tasso di evoluzione per un determinato taxon nel caso dei *branch models*.

Il confronto tra i modelli viene effettuato utilizzando un Likelihood Ratio Test (LRT):

$$\text{LRT} = 2\Delta l = 2(l_1 - l_0)$$

dove l è il logaritmo della likelihood di ogni modello. Il valore di LRT calcolato viene confrontato con una distribuzione di chi-quadrato (χ^2) in modo tale da determinare quale dei due modelli descriva meglio la situazione reale.

7.3 Analisi dei geni candidati

Il programma PAML è stato utilizzato in questo progetto per la stima di ω nelle sequenze codificanti dei geni candidati.

Ciascuna analisi è stata ripetuta 4 volte impostando 4 diversi valori di omega (0,1; 0,5; 1; 2) dai quali partire per l'esplorazione dello spazio dei parametri in modo da cercare di ridurre il rischio di ottenere dei valori di likelihood dovuti a massimi locali. Inoltre è stato imposto al programma di utilizzare un modello F3X4 per stimare π_j , ossia la frequenza relativa del codone j-esimo. Questo modello assume che la frequenza del codone dipenda dalla frequenza dei 4 possibili nucleotidi a ciascuna posizione del codone.

In particolare abbiamo effettuato:

1. un'analisi generale delle sequenze codificanti dei geni candidati in diverse specie per capire come varia omega, e quindi la pressione selettiva, lungo le sequenze (*site models*);
2. un confronto tra le sequenze di specie note per essersi adattate a vivere a basse pressioni di ossigeno e quelle di specie che vivono in condizioni di normossia (*branch models*).

La prima parte dello studio (*Site models*) ci ha permesso di capire se le sequenze codificanti dei geni candidati siano filogeneticamente conservate o se sono presenti delle regioni del gene che possano essere considerate punti chiave per l'azione della selezione naturale.

La seconda parte dello studio (*Branch models*) invece ha permesso di capire se durante la storia evolutiva di specie adattate a basse pressioni di ossigeno nell'aria, questi geni si siano evoluti con una velocità diversa da quella tipica delle altre specie. Se così fosse, specifiche varianti del gene potrebbero essersi fissate sotto la spinta della selezione diversificante nelle specie adattate ad ambienti in cui la pressione di ossigeno è ridotta.

7.3.1 Il database di sequenze

Per poter effettuare questi studi è necessario creare dati organizzare in un set di dati le sequenze codificanti dei geni candidati oggetto dello studio, provenienti da più specie.

Per fare ciò sono state consultate le banche dati pubbliche ENSEMBL Genome Browser (<http://www.ensembl.org/index.html>) ed NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>).

La ricerca è stata effettuata in modo da individuare il gene omologo di numerose specie con un alto livello di confidenza. In caso di sequenze ridondanti, ossia più sequenze depositate per lo stesso gene, sono state selezionate quelle aventi almeno una delle seguenti caratteristiche:

- la sequenza codificante completa (dal codone di start al codone di stop)
- il maggior numero di referenze
- la data di pubblicazione più recente.

7.3.2 CODEML

Il software PAML contiene al suo interno una serie di applicazioni specifiche per diverse tipologie di analisi. Noi abbiamo utilizzato l'applicazione CODEML, specifica per sequenze codificanti.

CODEML richiede l'allineamento delle sequenze relative al gene in esame, un albero filogenetico che rappresenti le relazioni tra le specie e un file di controllo dove vengono indicati vari parametri necessari per le iterazioni, tra cui i modelli da utilizzare per testare le ipotesi.

Le sequenze sono state allineate utilizzando il programma PRANKSTER (Loytynoja and Goldman 2005, Loytynoja and Goldman 2008) suggerito dallo stesso autore di PAML, in quanto contiene un algoritmo molto rigoroso in grado di ridurre gli errori di allineamento

(Fletcher and Yang 2010). Gli allineamenti sono stati effettuati a partire dalla sequenza proteica allo scopo di evitare errori dovuti allo slittamento della fase di lettura. Questo perché durante l'allineamento delle sequenze il programma introduce dei gap laddove siano presenti delle differenze tra le sequenze analizzate dovute a inserzioni o delezioni di basi, se si stanno allineando sequenze di DNA, o di amminoacidi, se si stanno allineando proteine.

Supponiamo di avere la seguente sequenza di DNA codificante (cDNA) e la corrispondente sequenza amminoacidica:

AATACCTCTTGTC

Asn Thr Ser Cys Pro (la cui abbreviazione è: N T S C P)

Supponiamo di allineare le sequenze di due specie e di ottenere i seguenti allineamenti:

Specie 1 AATACCTCTTGTC

Specie 2 AA-ACCT--TG-CCC

Nel momento in cui chiederemo al programma di convertire l'allineamento del cDNA in un allineamento di amminoacidi il programma automaticamente eliminerà i gap congiungendo fra loro le basi e provocando degli slittamenti della fase di lettura, che portano alla traduzione di una sequenza proteica diversa nelle due specie:

Specie 1 AAT ACC TCT TGT CCC → N T S C P

Specie 2 AAA CCT TGC CC → L P C

Per evitare di incorrere in questi errori si può seguire il procedimento inverso. Ossia una volta inserite nel programma le sequenze di cDNA queste vengono convertite in amminoacidi prima di effettuare l'allineamento rispettando così la fase di lettura di ciascuna sequenza. A questo punto si effettua l'allineamento delle sequenze amminoacidiche. Una volta effettuato l'allineamento si può chiedere al programma di visualizzare l'allineamento in forma di cDNA. Per fare questo il programma prende in considerazione la sequenza di cDNA di partenza e sfrutta le informazioni della fase di lettura imposta dall'allineamento proteico per risalire all'allineamento del cDNA di partenza.

Specie 1 N T S C P → AAT ACC TCT TGT CCC

Specie 2 - T - - T → --- ACC --- --- ACC

In questo modo i gap introdotti nell'allineamento della sequenza proteica saranno riportati nell'allineamento della sequenza di cDNA senza alterare la fase di lettura.

Non è stato possibile avere a disposizione lo stesso dataset di specie per tutti i geni perché i database pubblici non contengono le sequenze di tutti i geni per tutte le specie considerate. Per questo motivo in Appendice 3 e 4 sono indicate le specie utilizzate per ciascuna analisi con i codici identificativi del database pubblico dal quale sono state scaricate.

Le relazioni filogenetiche tra le specie sono state riassunte in alberi generati utilizzando il browser NCBI-Taxonomy (<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/>).

Abbiamo utilizzato questo browser perché come suggerito dall'autore del programma, quando possibile, gli alberi devono essere costruiti in modo da rappresentare le relazioni tassonomiche tra le specie.

Un'ulteriore verifica della correttezza degli alberi generati è stata effettuata utilizzando il database pubblico "Timetree" (<http://www.timetree.org/index.php>). Questa banca dati fornisce informazioni circa i tempi di divergenza tra le specie sfruttando i dati disponibili in letteratura. La ricerca viene effettuata indicando i nomi di due specie per la quali si desiderano avere queste informazioni, successivamente, attraverso un sistema gerarchico, vengono raccolte dalla letteratura tutte le informazioni disponibili circa i tempi di divergenza stimati in diversi lavori. I dati vengono poi riassunti in tabelle dove sono fornite tutte le informazioni riguardanti il tempo di divergenza (milioni di anni fa), il numero di geni utilizzati per la stima, il tipo di gene (ad esempio, nucleare o di organelli), tipo di dati (ad esempio nucleotidi e aminoacidi), e la fonte delle informazioni all'interno dell'articolo. Inoltre è possibile visualizzare anche i limiti di confidenza e gli errori standard (Hedges et al. 2006).

A questo punto avendo a disposizione le sequenze allineate e gli alberi filogenetici abbiamo applicato ai nostri dati i modelli implementati nell'applicazione CODEML.

7.3.3 Prima parte dello studio: Site Models

Nella prima fase di questo studio abbiamo effettuato un'analisi delle sequenze dei geni per capire se e come vari il valore di omega, e quindi la pressione selettiva, lungo la sequenza codificante dei geni candidati.

I *site models* utilizzati in questa prima fase dello studio considerano omega di ogni codone del gene come una variabile casuale di una distribuzione statistica, permettendo quindi ad omega di variare tra i diversi codoni.

Le relazioni tra le specie utilizzate sono rappresentate in figura 23.

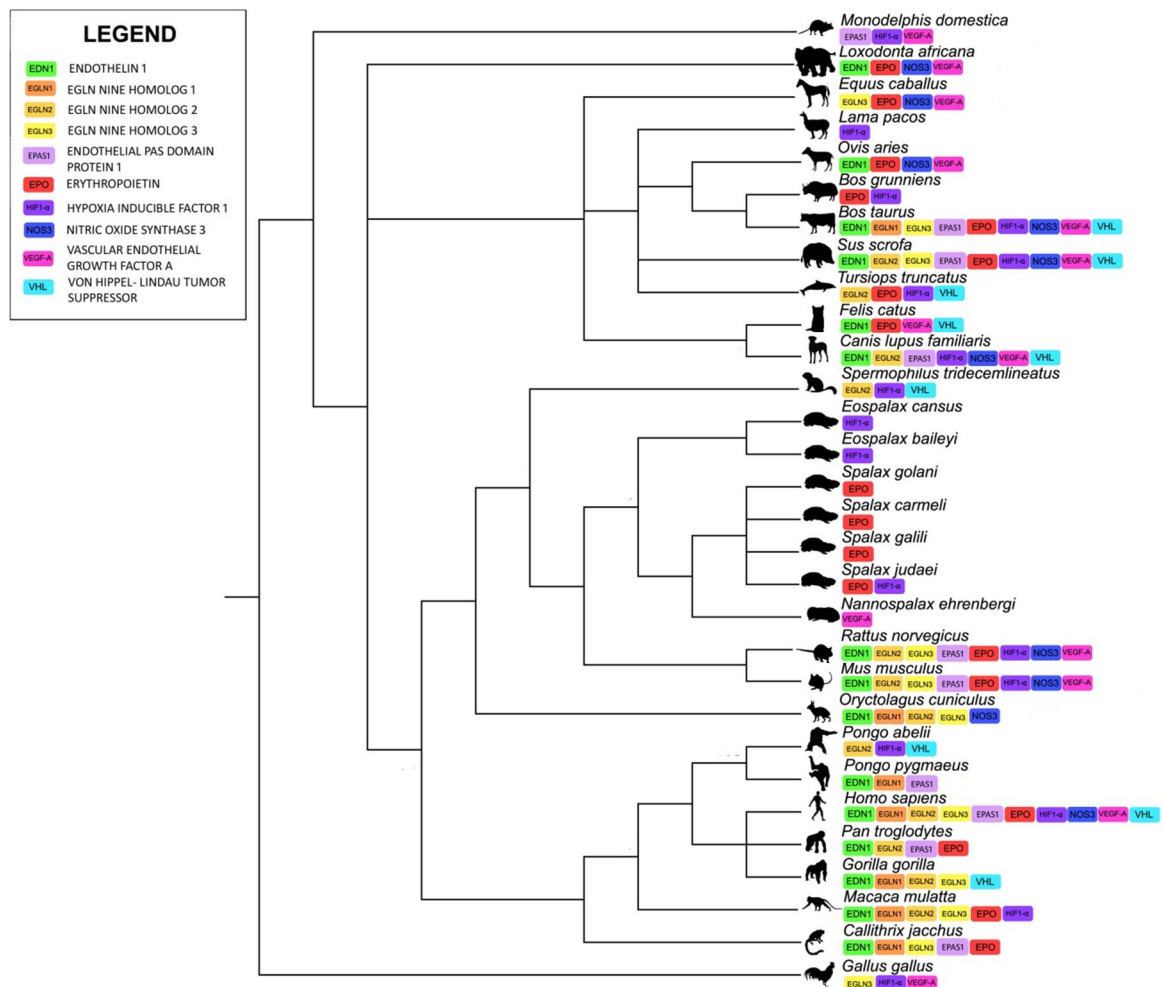


Fig.23 Relazioni tra le specie utilizzate per i *Site models* e rispettivi geni.

La prima domanda che ci siamo posti è stata:

- *Quali valori assume omega lungo la sequenza? Ci sono siti che assumono valori di omega compatibili con l'azione della selezione positiva ($\omega > 1$) e se si, quali?*

Per rispondere a queste domande sono stati utilizzati i modelli M1a e M2a ed i modelli M7 ed M8. I primi due sono modelli "discreti", conservativi, che rilevano la presenza di selezione positiva se questa agisce su una proporzione piuttosto elevata di siti. I modelli M7 ed M8 sono modelli "continui", ossia, ω non viene stimato per categorie definite da una proporzione di siti, ma si assume che ω sia distribuito in modo continuo secondo una specifica distribuzione. Questo permette di aumentare la risoluzione prendendo in considerazione più valori di omega lungo la sequenza e di individuare l'effetto della selezione positiva anche se questa agisce su un numero limitato di siti amminoacidici. Per questo motivo sono ritenuti modelli meno conservativi ((Yang 2006) pag274-275).

Modelli (M) applicati:

- *Ipotesi nulla (M1a, neutrale):* questo modello assume che lungo la sequenza esistano due classi di siti: una prima classe per la quale viene stimato il valore di ω assumendo che sia inferiore ad uno, ed una seconda classe per la quale il valore di ω viene fissato ad uno. Questo modello presenta due parametri stimati: il valore di $\hat{\omega}_0 < 1$ e la proporzione di siti con questo valore, \hat{p}_0 (gli altri due parametri sono uno fissato, $\omega_1=1$, mentre \hat{p}_1 è ricavabile come $1-\hat{p}_0$).
- *Ipotesi alternativa (M2a, selezione):* questo modello assume che lungo la sequenza esistano tre classi di ω : due classi comuni all'M1a, ed una terza in cui ω viene stimato assumendo sia superiore ad uno. Questo modello presenta quattro parametri stimati: $\hat{\omega}_0 < 1$ e la corrispondente proporzione di siti \hat{p}_0 , $\hat{\omega}_2 > 1$ e \hat{p}_1 , cioè la proporzione di siti con $\omega_1=1$ (gli altri due parametri sono $\omega_1=1$ fissato e la proporzione di siti con $\hat{\omega}_2 > 1$ ricavabile come $\hat{p}_2 = 1 - (\hat{p}_0 + \hat{p}_1)$).

La tabella 8 riassume i risultati dell'analisi M1a/M2a per i geni candidati di questo studio. I confronti tra i modelli mediante LRT non hanno dato risultati significativi (p-value > 0,05) nel nostro dataset di sequenze. Ciò favorisce l'ipotesi nulla secondo cui non ci sono siti che assumono valori di omega compatibili con la selezione positiva, ma le sequenze di

questi geni sono molto conservate come atteso in caso di neutralità o di selezione purificante previste dal modello M1a.

In particolare, la proporzione stimata di siti con valori di ω compresi tra 0,14 e 0,02 è in generale pari all'80-90%, con l'eccezione di EPO per cui scende al 76% e di EDN1 per cui viene stimata pari al 69%. Valori di ω così bassi sono indicativi di una forte selezione purificante che caratterizza gran parte delle sequenze proteiche codificate dai geni candidati di questo studio.

Tuttavia, come detto sopra, questi sono modelli discreti che suddividono la sequenza in categorie per le quali viene stimato un valore medio di omega sui siti che compongono la categoria stessa. Tale valore sarà ovviamente influenzato dal valore di omega assunto dalla maggior parte dei siti che compongono la categoria. Quindi se sono presenti molti siti con valori di ω minori o uguali ad 1 e pochi siti con valori di ω maggiori di 1, il valore medio di omega per quella categoria tenderà ad essere spostato verso valori negativi. Sarà quindi più difficile rilevare valori di $\omega > 1$ con questi modelli.

Modello	HIF1a (826 aminoacidi)		HIF2a (EPAS1) (870 aminoacidi)		EGLN1 (426 aminoacidi)		EGLN2 (407 aminoacidi)		EGLN3 (239 aminoacidi)	
	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri
M1a	-10013,46	$\hat{\rho}_0=0,92 \hat{\omega}=0,06$ ($\hat{\rho}_1=0,08$) $\omega_1=1$	-9708,23	$\hat{\rho}_0=0,90 \hat{\omega}=0,05$ ($\hat{\rho}_1=0,10$) $\omega_1=1$	-2589,18	$\hat{\rho}_0=0,80 \hat{\omega}=0,02$ ($\hat{\rho}_1=0,20$) $\omega_1=1$	-3677,77	$\hat{\rho}_0=0,94 \hat{\omega}=0,05$ ($\hat{\rho}_1=0,06$) $\omega_1=1$	-2245,24	$\hat{\rho}_0=0,94 \hat{\omega}=0,03$ ($\hat{\rho}_1=0,06$) $\omega_1=1$
M2a	-10013,46	$\hat{\rho}_0=0,92$ $\hat{\omega}_0=0,06$ $\hat{\rho}_1=0,08 \omega_1=1$ ($\hat{\rho}_2=0$) $\hat{\omega}_2=23,7\uparrow$	-9708,23	$\hat{\rho}_0=0,90 \hat{\omega}_0=0,05$ $\hat{\rho}_1=0,10 \omega_1=1$ ($\hat{\rho}_2=0$) $\hat{\omega}_2=63,2\uparrow$	-2588,92	$\hat{\rho}_0=0,86 \hat{\omega}_0=0,05$ $\hat{\rho}_1=0 \omega_1=1$ ($\hat{\rho}_2=0,14$) $\hat{\omega}_2=1,36$	-3677,77	$\hat{\rho}_0=0,94$ $\hat{\omega}_0=0,05$ $\hat{\rho}_1=0,06 \omega_1=1$ ($\hat{\rho}_2=0$) $\hat{\omega}_2=19,1\uparrow$	-2245,24	$\hat{\rho}_0=0,94 \hat{\omega}_0=0,03$ $\hat{\rho}_1=0,02 \omega_1=1$ ($\hat{\rho}_2=0,04$) $\hat{\omega}_2=1$
	LRT p value: 1		LRT p value: 1		LRT p value: 0,77		LRT p value: 1		LRT p value: 1	
M7	-9987,14	$\hat{\alpha}=0,31 \hat{\beta}=2,63$	-9667,97	$\hat{\alpha}=0,27 \hat{\beta}=2,15$	-2247,44	$\hat{\alpha}=0,01 \hat{\beta}=0,04$	-3675,61	$\hat{\alpha}=0,19 \hat{\beta}=1,80$	-2247,44	$\hat{\alpha}=0,14 \hat{\beta}=1,82$
M8	-9987,14	$\hat{\rho}_0=0,9997$ $\hat{\alpha}=0,31 \hat{\beta}=2,64$ ($\hat{\rho}_1=0,0003$) $\hat{\omega}_s=1$	-9667,97	$\hat{\rho}_0=0,99999 \hat{\alpha}=0,27$ $\hat{\beta}=2,15$ ($\hat{\rho}_1=1*10^{-5}$) $\hat{\omega}_s=1$	-2245,32	$\hat{\rho}_0=0,86 \hat{\alpha}=5,18$ $\hat{\beta}=99$ ($\hat{\rho}_1=0,14$) $\hat{\omega}_s=1,36$	-3675,61	$\hat{\rho}_0=0,99999$ $\hat{\alpha}=0,19 \hat{\beta}=0,18$ ($\hat{\rho}_1=1*10^{-5}$) $\hat{\omega}_s=1$	-2245,32	$\hat{\rho}_0=0,95 \hat{\alpha}=2,93$ $\hat{\beta}=99$ ($\hat{\rho}_1=0,05$) $\hat{\omega}_s=1$
	LRT p value: 1		LRT p value: 1		LRT p value: 0,71		LRT p value: 1		LRT p value: 0,12	

Modello	VHL (213 aminoacidi)		EPO (193 aminoacidi)		VEGFA (232 aminoacidi)		NOS3 (1203 aminoacidi)		EDN1 (212 aminoacidi)	
	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri	l	Stime dei parametri
M1a	-1776,14	$\hat{\rho}_0=0,84 \hat{\omega}=0,03$ ($\hat{\rho}_1=0,16$) $\omega_1=1$	-2837,75	$\hat{\rho}_0=0,76 \hat{\omega}=0,14$ ($\hat{\rho}_1=0,24$) $\omega_1=1$	-2263,88	$\hat{\rho}_0=0,82 \hat{\omega}=0,05$ ($\hat{\rho}_1=0,18$) $\omega_1=1$	-12108,17	$\hat{\rho}_0=0,96 \hat{\omega}=0,03$ ($\hat{\rho}_1=0,04$) $\omega_1=1$	-3338,22	$\hat{\rho}_0=0,69 \hat{\omega}=0,14$ ($\hat{\rho}_1=0,31$) $\omega_1=1$
M2a	-1776,14	$\hat{\rho}_0=0,84 \hat{\omega}_0=0,03$ $\hat{\rho}_1=0,10 \omega_1=1$ ($\hat{\rho}_2=0,06$) $\hat{\omega}_2=1$	-2837,75	$\hat{\rho}_0=0,76 \hat{\omega}_0=0,14$ $\hat{\rho}_1=0,14 \omega_1=1$ ($\hat{\rho}_2=0,10$) $\hat{\omega}_2=1$	-2263,88	$\hat{\rho}_0=0,82 \hat{\omega}_0=0,05$ $\hat{\rho}_1=0,18 \omega_1=1$ ($\hat{\rho}_2=0$) $\hat{\omega}_2=25,7\uparrow$	-12108,17	$\hat{\rho}_0=0,96 \hat{\omega}_0=0,03$ $\hat{\rho}_1=0,04 \omega_1=1$ ($\hat{\rho}_2=0$) $\hat{\omega}_2=1$	-3338,22	$\hat{\rho}_0=0,69 \hat{\omega}_0=0,14$ $\hat{\rho}_1=0,23 \omega_1=1$ ($\hat{\rho}_2=0,08$) $\hat{\omega}_2=1$
	LRT p value: 1		LRT p value: 1		LRT p value: 1		LRT p value: 1		LRT p value: 1	
M7	-1775,69	$\hat{\alpha}=0,08 \hat{\beta}=0,41$	-2837,31	$\hat{\alpha}=0,46 \hat{\beta}=0,98$	-2263,14	$\hat{\alpha}=0,18 \hat{\beta}=0,95$	-12093,88	$\hat{\alpha}=0,16 \hat{\beta}=2,52$	-3336,57	$\hat{\alpha}=0,44 \hat{\beta}=0,78$
M8	-1775,61	$\hat{\rho}_0=0,88 \hat{\alpha}=0,12$ $\hat{\beta}=1,92$ ($\hat{\rho}_1=0,12$) $\hat{\omega}_s=1$	-2836,49	$\hat{\rho}_0=0,88 \hat{\alpha}=0,77$ $\hat{\beta}=2,67$ ($\hat{\rho}_1=0,12$) $\hat{\omega}_s=1,12$	-2263,14	$\hat{\rho}_0=0,99999$ $\hat{\alpha}=0,18 \hat{\beta}=0,95$ ($\hat{\rho}_1=1*10^{-5}$) $\hat{\omega}_s=1$	-12087,28	$\hat{\rho}_0=0,99 \hat{\alpha}=0,20$ $\hat{\beta}=3,63$ ($\hat{\rho}_1=0,01$) $\hat{\omega}_s=1,75$	-3335,23	$\hat{\rho}_0=0,93 \hat{\alpha}=0,58$ $\hat{\beta}=1,32$ ($\hat{\rho}_1=0,07$) $\hat{\omega}_s=1,4$
	LRT p value: 0,92		LRT p value: 0,44		LRT p value: 1		LRT p value: 0,001		LRT p value: 0,26	

Tab.8 Valori di log-likelihood e stime dei parametri ottenute con i diversi modelli (*site models*) per i diversi geni analizzati.

l = logaritmo della likelihood

† in questi casi le analisi ripetute tendono a fornire diversi valori di ω ad indicare un risultato instabile dovuto al collasso del modello 2a nel modello 1a con meno parametri, come indicato da $\hat{\rho}_2=0$.

Applicando modelli "continui", meno conservativi, si può scendere maggiormente nel dettaglio.

Modelli (M) applicati:

- *Ipotesi nulla (M7, β , neutrale)*: questo modello assume che lungo la sequenza ω sia distribuito in modo continuo secondo una distribuzione β , definita tra 0 e 1. I parametri stimati sono due: $\hat{\alpha}$ e $\hat{\beta}$, che descrivono la forma della curva (Per maggior chiarezza in questa tesi i simboli α e β sono stati utilizzati al posto dei rispettivi p e q. utilizzati da PAML) (Fig.24).

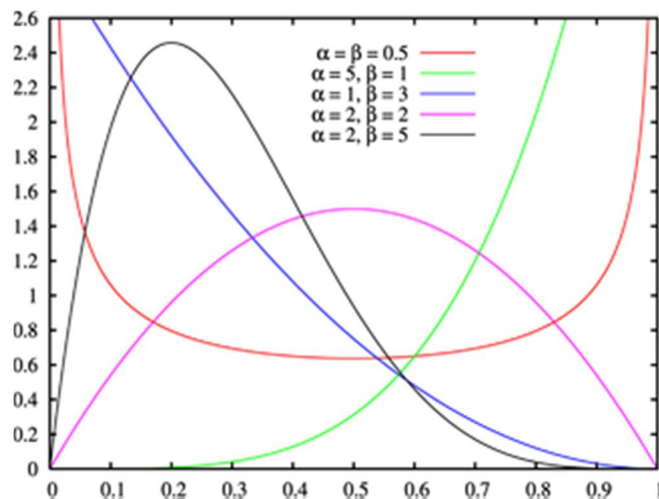


Fig.24 Distribuzione β . Sono indicate le diverse forme assunte dalla curva al variare di α e β .

- *Ipotesi alternativa (M8, β & ω , selezione)*: Questo modello contiene tutte le assunzioni del modello 7, in più aggiunge una classe siti per cui ω può assumere valori solo maggiori di 1 ($\hat{\omega}_s$). I parametri stimati sono $\hat{\alpha}$ e $\hat{\beta}$ della distribuzione β , la proporzione di siti con $\omega < 1$, \hat{p}_0 , ed $\hat{\omega}_s > 1$ (\hat{p}_1 , la proporzione di siti con valore $\hat{\omega}_s$ è ricavabile come $1 - \hat{p}_0$)

I risultati di questa analisi sono mostrati in tabella 8 per i geni candidati di questo studio.

Il confronto di questi modelli attraverso un LRT ha rilevato risultati significativi (p-value < 0,01) per il gene *NOS3*, mentre per gli altri geni i test non sono significativi confermando i

risultati dell'analisi precedente. In dettaglio, $\hat{\alpha}$ e $\hat{\beta}$ definiscono in tutti i casi una distribuzione β molto spostata verso bassi valori di ω .

Il risultato significativo riscontrato per il gene *NOS3* indica che alcuni siti nella sequenza del gene abbiano subito sostituzioni non-sinonime sotto la spinta della selezione adattativa e pertanto presentano valori di $\omega > 1$.

Se almeno uno dei modelli che testano la presenza di selezione positiva (M2a e M8) viene preferito al modello nullo, attraverso CODEML è possibile applicare un'analisi bayesiana che calcola la probabilità a posteriori di ciascun sito amminoacidico di appartenere ad una delle classi di omega individuate (Yang et al. 2005). L'approccio, descritto in dettaglio in Yang et al. (2005), è detto *bayes empirical bayes*, BEB, ed usa le stime di massima verosimiglianza dei parametri, quali la proporzione di siti di ogni classe di ω , associando a tali stime un'incertezza. Questo permette di distinguere i siti amminoacidici con valori di omega significativamente maggiori di 1 anche nel caso in cui i dataset non siano molto grandi.

NOS3 presenta 9 siti amminoacidici con una probabilità a posteriori maggiore del 50% di appartenere alla classe di siti sotto selezione positiva, di cui cinque hanno una probabilità superiore all'80%. La numerazione degli amminoacidi è riferita alla sequenza di *Homo sapiens* (Tab.9).

aa site	Prob.a post. ($\omega=1,7$)
36 Pro	0,55
118 Pro	0,78
235 Cys	0,92
304 Leu	0,98
786 Gly	0,80
824 Thr	0,88
872 Gln	0,61
1198 Ser	0,79
1201 Asn	0,85

Tab.9 Siti amminoacidici di *NOS3* con probabilità a posteriori maggiore del 50% di appartenere alla classe di $\omega > 1$. Il numero dell'amminoacido è riferito alla sequenza proteina dell'uomo (NCBI Reference Sequence: NP_000594.2).

Per capire dove sono localizzati questi amminoacidi nel contesto della struttura proteica abbiamo consultato il database UniProt, Universal Protein Resource (<http://www.uniprot.org/uniprot/P29474>), che contiene informazioni riguardanti la struttura e la funzione delle sequenze proteiche annotate. In particolare sulla base dei dati presenti in letteratura è in grado di fornire informazioni circa i domini proteici e gli amminoacidi che li compongono.

Sulla base delle informazioni presenti in questo database gli amminoacidi 118 P, 235 C e 304L fanno parte del dominio di interazione con la proteina NOSIP mentre gli amminoacidi 786G, 824T, 872Q sono localizzati in un dominio definito *FAD-binding-FR type*. Gli altri amminoacidi non sono associati a specifici domini proteici pertanto potrebbero svolgere esclusivamente un ruolo strutturale o potrebbero avere una funzione ancora non identificata.

Il dominio di legame FAD (flavin adenine dinucleotide) svolge un ruolo importante nei processi enzimatici che catalizzano reazioni biochimiche come deidrogenazione, trasferimento di elettroni, attivazione dell'O₂ in reazioni di ossidazione e idrossilazione. Si suppone che circa l' 1% di tutte le proteine di eucarioti e procarioti contengano un dominio di legame FAD (Mattevi 2006).

NOSIP (Nitric oxide synthase-interacting protein) è una proteina che regola la sintesi di Ossido Nitrico (NO) da parte di NOS3. In particolare l'interazione tra NOSIP e NOS3 favorisce la traslocazione di quest'ultimo dalla membrana plasmatica ai siti intracellulari inibendo la sintesi e il rilascio di NO (Dedio et al. 2001, Michel and Vanhoutte 2010). Tuttavia la regolazione della sintesi di NO non dipende solo da questa via di regolazione, ma si tratta di un sistema molto complesso che vede l'interazione di molti fattori (Michel and Vanhoutte 2010).

Queste informazioni indicano che alcuni siti amminoacidici che presentano una elevata probabilità a posteriori di essere sottoposti a selezione adattativa sono localizzati in domini che svolgono un ruolo funzionale all'interno della proteina.

Quindi per comprendere se tali sostituzioni possono modificare la struttura e la funzione della proteina in senso adattativo saranno necessari studi bioinformatici e di "molecular modeling".

Per quanto riguarda tutti gli altri geni questa prima parte dello studio ci permette di concludere che le sequenze amminoacidiche delle proteine codificate dai geni candidati sono generalmente conservate grazie a pressioni selettive negative ($0 < \omega < 1$) che tendono a mantenerne invariata la sequenza durante l'evoluzione. Questo è compatibile con la teoria neutrale secondo cui la maggior parte dei geni evolve sotto l'azione della selezione purificante o in condizioni di neutralità.

Le proteine codificate dai geni candidati analizzati in questo studio sono coinvolte nella regolazione di un ciclo metabolico complesso e nella maggior parte dei casi non sono associate alla manifestazione diretta di un fenotipo, ma piuttosto questo dipende dall'interazione di più fattori. Per questo è possibile che la loro evoluzione sia più lenta.

È anche possibile ipotizzare che la selezione positiva abbia agito in passato portando a fissazione alcune varianti che favoriscono la corretta modulazione di questo ciclo metabolico e che oggi noi non rileviamo appunto perché fissate in tutte le specie. Una delle conseguenze di questo fenomeno è appunto un incremento della frequenza delle varianti neutrali nelle regioni circostanti il sito selezionato (*complete selective sweep*).

Nel contesto di questa tesi questi risultati sono stati presi in considerazione nella scelta delle regioni geniche da sequenziare nel nostro campione. In particolare, considerata la notevole conservazione delle regioni codificanti, l'analisi ci ha suggerito di aggiungere al nostro studio il sequenziamento di importanti regioni di controllo.

7.3.4 Seconda parte dello studio: Branch models

I geni candidati presi in considerazione codificano per proteine coinvolte in numerosi processi fisiologici, non solo legati alla risposta cellulare all'ipossia. Le analisi precedenti ci hanno dato informazioni su come si sono evolute queste sequenze proteiche tuttavia non sono stati presi in considerazione fattori legati all'adattamento all'ipossia.

In questa seconda parte dello studio sono stati applicati dei modelli in grado di valutare la presenza di differenze nei tassi di sostituzione tra specie e alcune delle quali caratterizzate dall'adattamento ad ambienti ipossici..

Confrontare due linee distinte da un particolare adattamento permette di ipotizzare che eventuali differenze nei tassi di sostituzione siano dovute alla caratteristica che le contraddistingue, in questo caso l'ipossia. Questo permette di individuare più facilmente casi di selezione adattativa, diversificante.

Le analisi effettuate con questi modelli sono ristrette ai geni *EPO* ed *HIF1a* poiché tra i geni candidati sono gli unici per i quali sono disponibili le sequenze per la maggior parte delle specie di mammiferi che vivono in condizioni di ipossia considerate in questa parte dello studio.

7.3.4.1 Il dataset

Le specie di mammiferi adattate ad ambienti "ipossici" per le quali sono disponibili le sequenze di almeno uno dei due geni sopra citati sono: *Vicugna pacos* (Alpaca), *Bos grunniens* (Yak o bue tibetano), *Pantholops hodgsonii* (Antilope tibetana), alcune specie del genere *Spalax* (topo-talpa), *Tursiops truncatus* (delfino).

L'Alpaca appartiene alla famiglia dei *Camelidae* e rappresenta una delle specie domestiche più diffuse in Sud America insieme al Lama (*Lama glama*). Vive ad altitudini comprese tra i 3500 e i 5000 metri sulle Ande del Perù meridionale, della Bolivia settentrionale e del nord del Cile. Lo Yak, o bue tibetano, e l'Antilope tibetana appartengono entrambi alla famiglia dei *Bovidae* e sono diffusi negli altopiani del Tibet dove vivono a quote anche superiori ai 5000m.

Spalax rappresenta una specie di ratto-talpa cieca appartenente alla famiglia degli Spalacidae, che vive sotto terra a concentrazioni di O₂ e CO₂ variabili. Rappresenta il mammifero in grado di tollerare meglio e per tempi più lunghi l'ipossia: test di laboratorio hanno dimostrato che ha la capacità di sopravvivere anche per 14 ore a concentrazioni di ossigeno pari al 3% rispetto al ratto che decede dopo sole 2-4 ore (Shams et al. 2004).

Tursiops truncatus, più comunemente noto come delfino dal naso a bottiglia è un cetaceo odontoceto appartenente alla famiglia dei Delfinidi. I cetacei sono comunemente noti per essere mammiferi completamente adattati alla vita acquatica e in grado di rimanere sott'acqua senza respirare (ipossia) per periodi di tempo molto più lunghi di tutti gli altri mammiferi.

Le specie utilizzate nell'applicazione di questi modelli sono indicate in Appendice 4 e le relazioni tra le specie utilizzate sono rappresentate in figura 25.

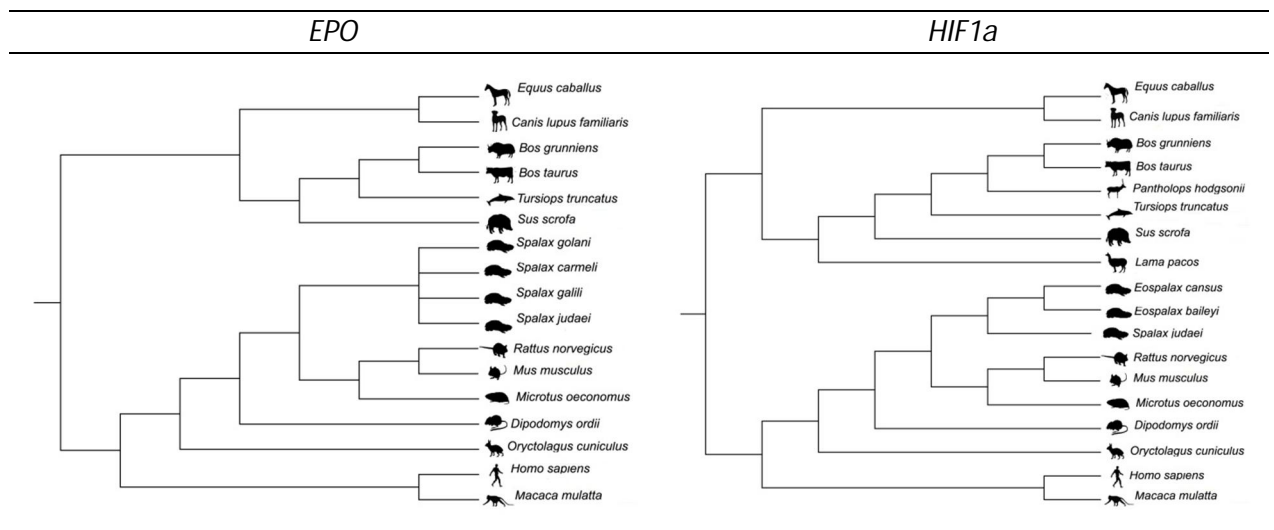


Fig.25 Relazioni tra le specie utilizzate per i *Branch Models* per i geni *EPO* e *HIF1a*.

7.3.4.2 Analisi dell'insieme delle specie adattate all'ipossia

Anche in questo caso abbiamo fatto delle ipotesi a cui abbiamo cercato di dare una risposta sfruttando i modelli di CODEML che stimano come varia ω tra i rami della filogenesi. In questi modelli è necessario specificare qual è la linea filogenetica per la quale si vuole testare l'ipotesi di un tasso evolutivo distinto dal resto dell'albero. Questi rami della filogenesi saranno segnalati al programma come rami *foreground*, mentre tutti gli altri saranno considerati rami *background*.

Nel nostro caso sono definiti *foreground* i rami della filogenesi che portano a specie adattate all'ipossia e *background* i rami che portano a specie che vivono in condizioni normali di ossigeno.

La prima cosa ci siamo chiesti è stata:

- *Le specie adattate all'ipossia e le specie che vivono in normali condizioni di ossigeno hanno subito pressioni selettive differenti? Ossia, i rami foreground assumono valori di omega differenti da quelli dei rami background?*

Per rispondere a questa domanda abbiamo applicato i modelli più semplici definiti *branch models* che permettono ad ω di variare lungo i rami della filogenesi, ma non tra i diversi siti, stimando un ω medio su tutti i siti (Yang 1998, Yang 2007b). Questo permette di testare se ci sono differenze nei tassi di sostituzione tra i rami della filogenesi definiti come descritto sopra.

Modelli (M) applicati:

- *Ipotesi nulla (one ratio):* ω assume lo stesso valore su tutti i rami.
- *Ipotesi alternativa (two-ratio):* il/i ramo/i *foreground* assumono un valore di ω diverso rispetto ai rami *background*, stimato dal programma.

Se il confronto tra i modelli mediante LRT favorisce il modello alternativo allora si può ipotizzare che le specie adattate all'ipossia si siano evolute con un tasso differente rispetto alle altre specie.

Una prima analisi è stata svolta raggruppando sotto la definizione di *foreground* i rami di tutte le specie adattate all'ipossia. Il confronto dei modelli in questo caso non ha portato risultati significativi né per EPO né per HIF1a ($p\text{-value}>0,05$) indicando che non ci sono sufficienti evidenze per affermare che le specie adattate all'ipossia abbiano valori di omega significativamente differenti da quelli di tutte le altre specie per questi due geni.

Tuttavia le specie raggruppate come *foreground* sono adattate a condizioni diverse di ipossia e da tempi differenti. Se avessimo rilevato un diverso tasso evolutivo per tutte le specie ipossiche, pur molto distanti evolutivamente, avremmo concluso che l'adattamento sia stato raggiunto attraverso una forte convergenza che però, dato il numero di fattori implicati nella risposta fisiologica all'ipossia, potrebbe non essere molto probabile. Per questo abbiamo ripetuto le analisi considerando le specie adattate all'ipossia una alla volta come ramo *foreground*, escludendo dall'analisi le altre specie ipossiche.

7.3.4.3 Analisi delle singole specie adattate all'ipossia

Questa analisi ci ha permesso di individuare un valore specifico di omega per ognuno dei rami *foreground* testati nelle diverse analisi. Per quanto riguarda le specie di Spalax, queste sono state considerate tutte insieme indicando come ramo *foreground* quello che porta alla famiglia degli Spalacidae.

I risultati di questa analisi sono mostrati in tabella 10.

Per quanto riguarda il gene *EPO* nessuno dei confronti è risultato significativo ($p\text{-value}>0,05$) indicando che le specie che vivono in normali condizioni di normossia e le specie adattate all'ipossia, anche quando considerate separatamente, non mostrano differenze significative nei tassi di sostituzione.

Se si considera invece il gene *HIF1a*, il confronto mediante LRT tra il modello nullo e il modello alternativo ha mostrato risultati significativi nel test in cui come specie adattata all'ipossia è stato considerato il *Bos grunniens* (Yak). Sebbene il valore stimato di ω sia 0,89 per lo Yak e dunque non sia indicativo di selezione positiva per l'intero gene, la significatività del test ci indica che questa linea abbia subito un'accelerazione evolutiva

rispetto al resto dell'albero che potrebbe essere il risultato di un adattamento a condizioni di bassa pressione di ossigeno.

	Modello nullo	Modello alternativo		
Specie <i>foreground</i>	ω	ω (ipossia)	ω (normossia)	LRT p-value
<i>EPO</i>				
<i>Bos grunniens</i>	0,24	0,40	0,23	1
Spalacidae	0,24	0,21	0,25	0,546
<i>Tursiops truncatus</i>	0,25	0,32	0,24	0,613
<i>HIF1a</i>				
<i>Bos grunniens</i>	0,10	0,89	0,10	0,003*
<i>Vicugna pacos</i>	0,10	0,14	0,10	0,404
<i>Tursiops truncatus</i>	0,09	0,04	0,10	0,122
Spalacidae	0,09	0,17	0,08	0,113
<i>Pantholops hodgsonii</i>	0,10	0,04	0,10	0,239

Tab.10 Branch models.

Nella prima colonna sono indicate per ciascuno dei geni le specie adattate all'ipossia i cui rami sono stati marcati come *foreground*. In tabella sono riportati i valori di omega (dN/dS) stimati dal programma per il ramo *foreground* in esame e per tutti gli altri rami raggruppati come *background*. Il risultato del LRT è riferito al confronto tra il modello nullo che ipotizza lo stesso valore di omega su tutti i rami e il modello alternativo che ipotizza valori diversi di omega nei rami *background* e *foreground*. All'interno della famiglia degli Spalacidae rientrano più specie di Spalax elencate in appendice 4 e considerate tutte insieme nelle analisi.

* p-value < 0,05

8. VARIABILITÀ GENETICA DEI GENI CANDIDATI *EPO*, *VHL* E *EPAS1 (HIF-2 α)* E ASSOCIAZIONE CON L'ALTA QUOTA

I geni candidati selezionati mediante gli approcci descritti nel capitolo 6 fanno parte della via metabolica di HIF e comprendono sia geni la cui regolazione dipende direttamente da HIF sia proteine che interagiscono con esso per regolarne l'attività.

Tra questi abbiamo selezionato *EPO* ed *VHL* come geni da cui partire per il sequenziamento dei geni candidati per i motivi spiegati nell'introduzione. Il primo è un gene la cui espressione è regolata da HIF e per il quale abbiamo a disposizione le misurazioni, effettuate nei nostri campioni, del fenotipo maggiormente associato all'espressione del gene, cioè il valore del contenuto di emoglobina nell'unità di volume del sangue. Il secondo codifica per una proteina che favorisce il mantenimento di livelli normali di HIF inducendone la degradazione in condizioni di normossia ed è stato spesso trovato associato a patologie legate a condizioni di ipossia, come determinati tipi di tumore.

I risultati dei sequenziamenti hanno permesso di:

1. stimare indici di variabilità genetica intra-popolazione e misure di differenze tra popolazioni in grado di fornire informazioni riguardanti l'evoluzione del locus in esame in popolazioni umane adattate all'ipossia;
2. verificare se esistono differenze tra popolazioni di alta quota (nostri campioni) e popolazioni di bassa quota (Maya HGDP) per i suddetti polimorfismi;
3. testare l'eventuale presenza di associazione tra i polimorfismi rilevati e i fenotipi misurati nel nostro campione.

8.1 *EPO*

Per verificare se i risultati delle analisi interspecifiche, che hanno dimostrato che le regioni codificanti dei geni candidati sono generalmente molto conservate, fossero estendibili anche alla variabilità all'interno delle nostre popolazioni, abbiamo effettuato una prima analisi delle regioni codificanti e regolatrici del gene *EPO* in un sottogruppo di individui (19 individui, 38 cromosomi). Questo ci ha permesso di indirizzare i sequenziamenti successivi sulle regioni maggiormente polimorfiche.

Nello specifico sono state amplificate e sequenziate 1322 bp della regione codificante, comprendente gli esoni 2, 3 e 4. Per quanto riguarda le regioni regolatrici sono state amplificate e sequenziate 493bp della regione del promotore e 502bp comprendenti tutta la sequenza dell'elemento 3' Hypoxia Responsive Element (HRE: 259bp), per un totale di 995 bp (Fig.10 paragrafo 4.1.2.2).

Le analisi dei cromatogrammi prodotti dal sequenziamento hanno permesso di rilevare la mancanza di variabilità nelle regioni codificanti e la presenza di due polimorfismi nelle regioni regolatrici. I polimorfismi identificati sono presenti nel database dbSNP (Entrez SNP <http://www.ncbi.nlm.nih.gov/projects/SNP/>) con gli identificativi *rs1617640* (A/C) e *rs551238* (G/T) (Tab.11).

rs	Posizione cromosomica*	Regione del gene	Allele ancestrale	Allele derivato
rs1617640	100317298	Promotore	A	C
rs551238	100321528	HRE	G	T
Distanza: 4230 bp				

Tab.11 Siti polimorfici di *EPO* identificati nel nostro studio.

*Genome build release 37.1 GRCh37, elica forward

Il polimorfismo *rs1617640* è localizzato nella regione del promotore 1306bp a monte dell'ATG di inizio trascrizione. Nello studio di Tong et al. (2008) la variante ancestrale di tale polimorfismo è stata associata ad una maggiore espressione di *EPO* in cellule renali probabilmente a causa della formazione di un sito di legame per proteine con attività di *enhancer* in grado di stimolare una maggiore trascrizione del gene (Tong et al. 2008).

Il polimorfismo *rs551238* localizzato nella regione HRE, 79 bp a valle della sequenza consenso di legame per HIF2a, è indicato anch'esso nel lavoro di Tong et al. (2008) in quanto presenta completo linkage disequilibrium con il polimorfismo sopra descritto.

Allo scopo di analizzare la variabilità genetica riscontrata dalle analisi precedenti in un campione più ampio sono state amplificate e sequenziate le regioni del promotore e di HRE in un totale di 89 individui (178 cromosomi) così suddivisi:

- 16 individui provenienti dal sito di campionamento di Cusibamba,
- 28 individui dal sito di Cusco,
- 45 individui campionati a Puno.

Le stesse regioni sono state sequenziate anche in 25 individui Maya di HGDP scelti come popolazione di controllo di bassa quota come descritto nel paragrafo 3.2.

8.1.1 *EPO*: variabilità genetica all'interno e tra popolazioni di alta quota

I risultati del sequenziamento hanno confermato la presenza dei due polimorfismi sopra descritti sia nel campione andino di alta quota sia nel campione di bassa quota rappresentato da individui Maya. L'analisi delle sequenze non ha rilevato la presenza di altri polimorfismi nei suddetti campioni.

La determinazione della fase aplo-tipica e la stima del linkage disequilibrium hanno permesso di concludere che sia nelle popolazioni andine che nei Maya gli SNP *rs1617640* e *rs551238* mostrano completo LD ($D'=1$) e di stimare inoltre le frequenze aplo-tipiche e genotipiche nei diversi campioni (Tab.12).

		AT	CG	AT/AT	AT/CG	CG/CG
Andini	Femmine (n: 55)	0,88	0,12	0,80	0,16	0,04
	Maschi (n: 34)	0,84	0,16	0,71	0,26	0,03
Cusco	Femmine (n: 19)	0,95	0,05	0,89	0,11	0,00
	Maschi (n: 9)	0,89	0,11	0,78	0,22	0,00
Cusibamba	Femmine (n: 10)	0,95	0,05	0,90	0,10	0,00
	Maschi (n: 6)	0,75	0,25	0,50	0,50	0,00
Puno	Femmine (n: 26)	0,81	0,19	0,69	0,16	0,04
	Maschi (n: 19)	0,84	0,16	0,74	0,21	0,05
Maya	n: 25	0,86	0,14	0,72	0,28	0,00

Tab.12 *EPO*: frequenze aplo-tipiche e genotipiche e in maschi e femmine di alta quota. Mettere i valori nelle singole popolazioni

Le popolazioni sono in equilibrio di Hardy-Weinberg per questo locus. Le misure degli indici di variabilità genetica sono riportate in tabella 13.

<u>EPO</u>	CS+CB+PN	Cusco (CS)	Cusibamba (CB)	Puno (PN)	Maya
N. di cromosomi	178	56	32	90	50
N. di siti	995	995	995	995	995
N. di siti polimorfici	2	2	2	2	2
N. di aplotipi	2	2	2	2	2
Diversità aplotipica $H \pm SD$	$0,24 \pm 0,04$	$0,14 \pm 0,06$	$0,23 \pm 0,09$	$0,30 \pm 0,05$	$0,25 \pm 0,07$
$\pi \pm SD (x10^{-3})$	$0,47 \pm 0,08$	$0,27 \pm 0,12$	$0,45 \pm 0,18$	$0,59 \pm 0,10$	$0,49 \pm 0,14$
$\theta_s \pm SD$ (per sites)	$0,35 \pm 0,25$	$0,43 \pm 0,32$	$0,50 \pm 0,37$	$0,39 \pm 0,29$	$0,45 \pm 0,33$

Tab.13 *EPO*: Indici di variabilità genetica intra-popolazione.

Tra le popolazioni di alta quota il campione di Cusco mostra livelli di diversità aplotipica e diversità nucleotidica inferiori rispetto alle altre due popolazioni. Tuttavia i tre gruppi andini risultano geneticamente poco differenziati, come mostrato dai valori di F_{ST} in tutti i casi inferiori al 5% e non significativi (Tab.14).

	Cusco	Cusibamba	Puno	Maya
Cusco	0			
Cusibamba	0	0		
Puno	0,033	0	0	
Maya	0,006	0	0	0
Andini ⁺	-	-	-	0

Tab.14 *EPO*: F_{ST} a coppie

⁺ Andini: si intendono le popolazioni di Cusco, Cusibamba e Puno considerate tutte insieme.

8.1.2 *EPO*: confronti alta e bassa quota

La tabella 14 mostra come le popolazioni andine e il campione Maya non mostrino alcuna divergenza per quanto riguarda i polimorfismi rilevati nelle regioni fiancheggianti il gene *EPO*. Questo induce ad ipotizzare che non esistano vantaggi in termini di adattamento all'alta quota legati ad una particolare combinazione di varianti alleliche a questo locus.

8.1.3 *EPO*: associazione genotipo/fenotipo

Per verificare questa ipotesi, pur non avendo a disposizione i dati fenotipici relativi ai campioni di bassa quota, è possibile effettuare delle analisi di associazione genotipo/fenotipo all'interno del campione di alta quota, prendendo in considerazione sia tutti gli individui del campione, sia considerando solo gli individui che manifestano valori estremi delle variabili fenotipiche.

Tutte le analisi sono state effettuate separando gli individui in base al sesso. Questa suddivisione risulta necessaria poiché i fenotipi misurati presentano valori fisiologicamente differenti tra maschi e femmine.

I fenotipi sono rappresentati dalle seguenti variabili quantitative continue: indice di massa corporea (BMI), pressione sistolica (Sis) e pressione diastolica (Dia) espresse in millimetri di mercurio (mmHg), frequenza del battito cardiaco espressa in battiti per minuto (BPM), percentuale di saturazione di ossigeno (SpO₂), temperatura corporea espressa in °C (T°), emoglobina espressa in mg/dL (Hb), percentuale di ematocrito (HTC), livelli di glucosio espressi in mg/dL (Glu) e misurati solo per il campione di Puno.

Il campione in esame è costituito da 55 femmine e 34 maschi. Le frequenze aplotipiche e genotipiche per sesso e per popolazione sono riportate in tabella 12.

Età ed etnia (Quechua/Aymarà) si distribuiscono in modo omogeneo nei tre gruppi genotipici (ANOVA e Test esatto di Fisher tabella di contingenza 3x2).

Questo permette di escludere un eventuale effetto confondente dovuto a queste variabili (stratificazione) sui nostri test di associazione.

Per testare l'associazione tra i tre genotipi (gruppi) e i fenotipi sopra descritti è stata utilizzata l'analisi della varianza (ANOVA). Questa analisi permette di verificare se il valore medio della variabile fenotipica in esame presenta differenze nelle categorie determinate dai genotipi. Se così fosse è possibile che uno dei gruppi, corrispondente ad un determinato genotipo, presenti un valore medio della variabili fenotipica che differisce da quello degli altri gruppi (genotipi).

I risultati di queste analisi hanno mostrato assenza di associazione tra i genotipi al locus *EPO* e tutti i fenotipi analizzati, inclusi i valori di ematocrito (Fig.26), fenotipo direttamente correlato all'espressione del gene.

L'analisi dei fenotipi estremi descritta nel paragrafo 4.2.5.1, è stata utilizzata per verificare l'ipotesi nulla che gli individui appartenenti alle tre categorie genotipiche (AT/AT; CG/CG; AT/CG) si distribuiscano in modo indipendente nelle due categorie fenotipiche estreme e che pertanto non ci sia associazione tra un dato genotipo e l'appartenenza ad una delle due categorie fenotipiche.

I test effettuati per ciascuna variabile fenotipica (Test esatto di Fisher, tabella di contingenza 3x2) non sono risultati significativi sottolineando ancora una volta l'indipendenza genotipo/fenotipo.

In conclusione nonostante questi polimorfismi siano localizzati in regioni regolatrici dell'espressione genica non sembrano influenzare le variabili fenotipiche misurate nel campione di alta quota.

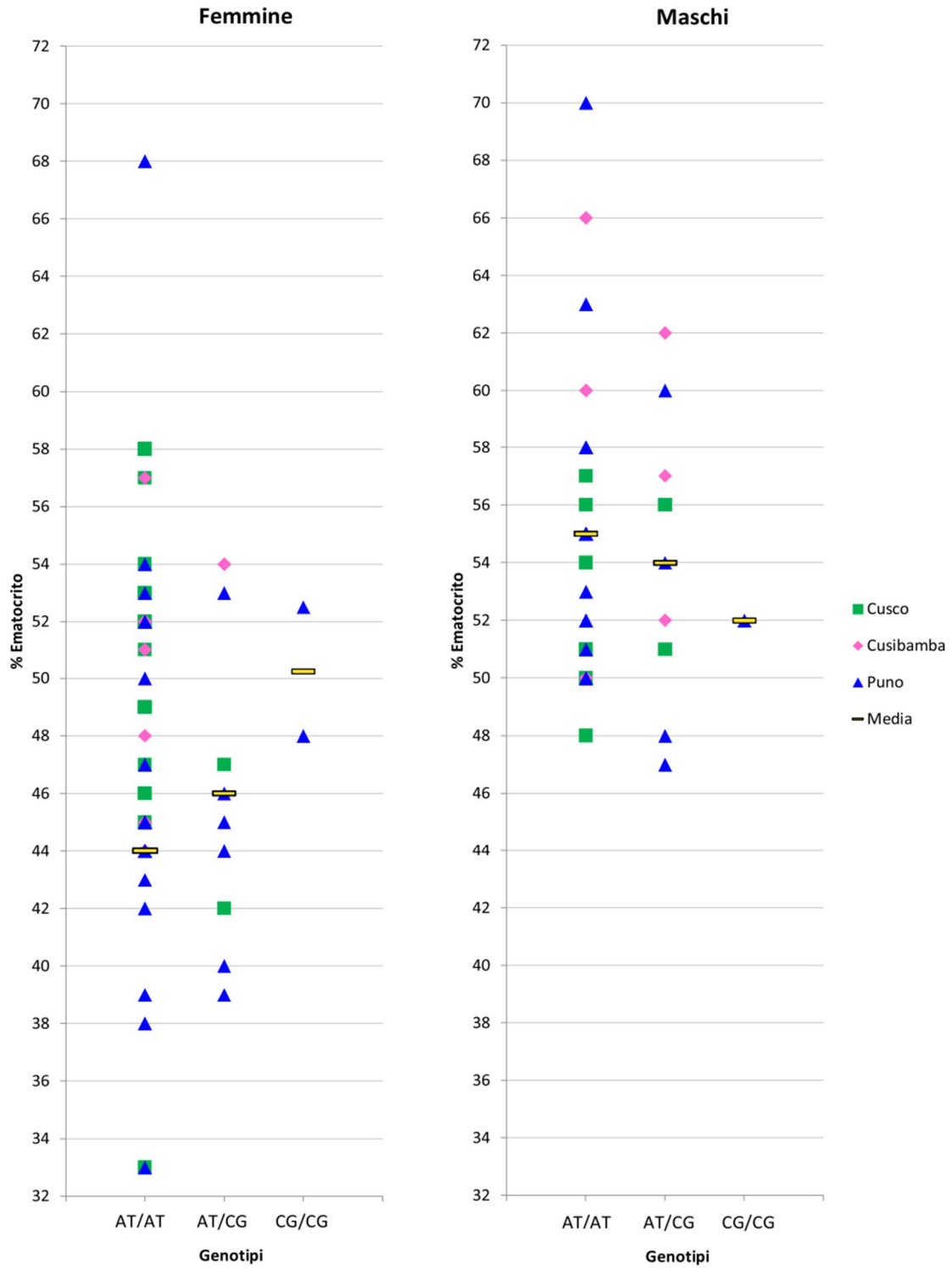


Fig.26 Valori di ematocrito in maschi e femmine di alta quota in relazione al genotipo al locus *EPO*.

8.2 VHL

La strategia di sequenziamento utilizzata per VHL corrisponde a quella utilizzata per il gene EPO, pertanto abbiamo effettuato un primo screening dell'intera sequenza del gene comprendente sia le regioni codificanti che le regioni regolatrici non tradotte (UTR) in un campione di 11 individui (22 cromosomi), prima di procedere all'analisi della variabilità genetica in un campione più ampio. Le regioni analizzate prodotte dall'amplificazione e sequenziamento del gene sono indicate in figura 11 paragrafo 4.1.2.3. Nello specifico sono state amplificate e sequenziati 4 frammenti, un primo frammento comprendente la regione 5'UTR e l'esone 1, un secondo frammento comprendente l'esone 2 e un terzo frammento comprendente l'esone 3 e parte della regione 3'UTR.

Anche in questo caso i risultati del sequenziamento nelle nostre popolazioni concordano con le analisi di evoluzione molecolare a livello interspecifico indicando che la regione codificante di questo gene è complessivamente monomorfica ed evolutivamente conservata.

Ancora una volta le regioni regolative del gene, non tradotte (UTR), mostrano un maggiore livello di variabilità dato dalla presenza di 6 siti polimorfici, noti nel database dbSNP (Entrez SNP <http://www.ncbi.nlm.nih.gov/projects/SNP/>) con gli identificativi riportati in tabella 15.

rs	Posizione cromosomica*	Distanza in bp dal polimorfismo precedente	Regione del gene	Allele ancestrale	Allele derivato
rs779804	10183274		46 bp a monte di 5'UTR	T	C
rs779805	10183337	63	5'UTR	G	A
rs1642742	10191943	8606	3'UTR	G	A
rs1681669	10192672	729	3'UTR	G	A
rs1681668	10192709	37	3'UTR	C	T
rs1642741	10192762	53	3'UTR	C	T

Tab.15 Siti polimorfici di VHL identificati nel nostro studio.

*Genome build release 37.1 GRCh37, elica forward.

Nessuno di questi polimorfismi risulta associato a fenotipi patologici in letteratura, tuttavia nel lavoro di Liu et al (2004) la variante A del polimorfismo *rs779805* (G>A) risulta in LD con l'allele T del polimorfismo associato all'insorgenza della policitemia di Chuvash

(rs28940298 C>T)(Liu et al. 2005). Nei nostri campioni tale polimorfismo è assente e tutti gli individui presentano la variante C in accordo con lo stato di salute di tutti i pazienti. Inoltre il polimorfismo rs779805, riscontrato nei nostri campioni, negli studi di Banks et al. (2006-2009) è stato correlato, seppur in modo non significativo, ad una maggiore metilazione del gene in individui malati di carcinoma delle cellule renali, tuttavia questi risultati non sono stati approfonditi in studi successivi (Banks et al. 2006, Young et al. 2009).

Quindi, considerati i risultati di questo primo sequenziamento e data l'importanza degli esoni 1 e 3 nella sintesi dei domini fondamentali della proteina abbiamo indirizzato le successive analisi su queste due regioni escludendo dai sequenziamenti il frammento monomorfo comprendente l'esone 2. Le analisi sono state estese al campione di 89 individui (178 cromosomi) già tipizzati per il gene *EPO* e al campione di controllo rappresentato dai 25 individui Maya di bassa quota.

I risultati del sequenziamento di 178 cromosomi hanno rilevato la presenza di 6 siti polimorfici, riportati in tabella 15. Anche nel campione più ampio la variante polimorfica associata a policitemia di Chuvash è assente.

8.2.1 *VHL*: variabilità genetica all'interno e tra popolazioni di alta quota

I polimorfismi in tabella 15 individuati nel campione Andino sono stati rilevati anche nel campione di bassa quota, Maya, seppure con frequenze differenti (Tab.16).

Aplotipi	Andini		CUSCO		CUSIBAMBA		PUNO		MAYA	
	n	Freq.	n	Freq.	n	Freq.	n	Freq.	n	Freq.
H1 T A A A T C	77	0,433	27	0,482	9	0,282	41	0,456	31	0,620
H2 T G G G C C	44	0,247	11	0,196	11	0,344	22	0,244	13	0,260
H3 C G G G C T	53	0,298	16	0,286	10	0,312	27	0,300	5	0,100
H4 C G G G C C	1	0,006	0	0	1	0,031	0	0	0	0
H5 C G A A T C	1	0,006	0	0	1	0,031	0	0	0	0
H6 T A G G C T	2	0,010	2	0,036	0	0	0	0	1	0,020
Tot	178	1,000	56	1,000	32	1,000	90	1,000	50	1,000

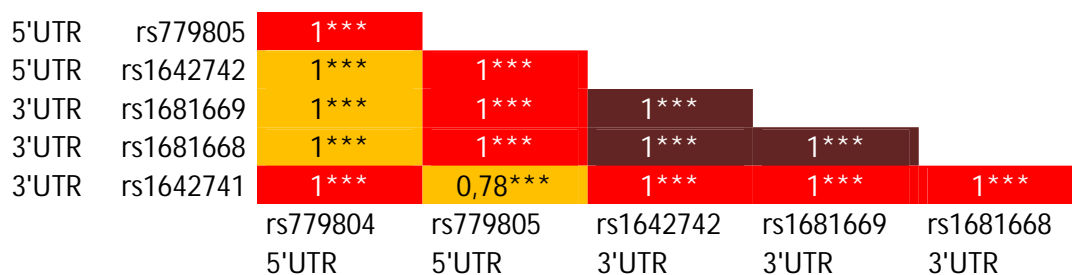
Tab.16 *VHL*: aplotipi e relative frequenze nel campione di alta quota (Andini) e nei campioni di bassa quota (Maya).

Il linkage disequilibrium tra i 6 SNP di *VHL* è stato testato con due indici, il D' e $L'R^2$ e i risultati sono sintetizzati in figura 27.

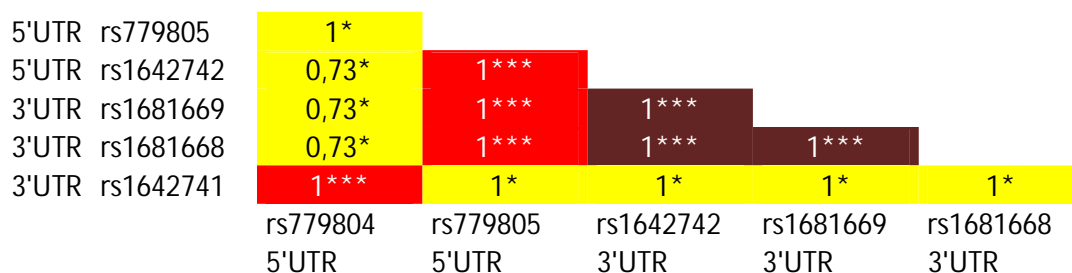
R^2 : legenda



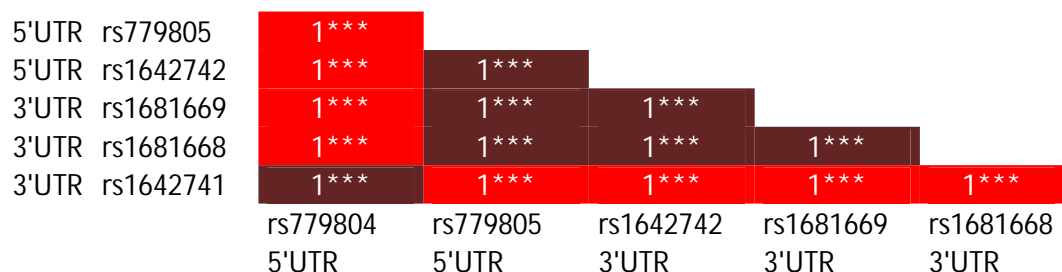
Cusco



Cusibamba



Puno



Maya

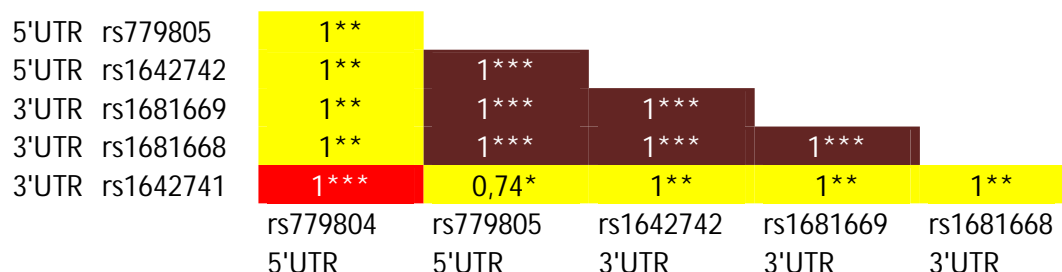


Fig.27 Rappresentazione schematica dei valori di linkage disequilibrium tra i 6 SNP delle regioni 5' e 3' UTR di *VHL* nelle popolazioni di alta quota (Cusco, Cusibamba e Puno) e nei Maya di bassa quota. La matrice triangolare mostra i valori di $|D'|$ tra coppie di SNP e i colori rappresentano i valori di R^2 . La significatività dell'associazione tra due loci è stata calcolata per ogni confronto attraverso un test esatto di Fisher (* $P < 0,05$; ** $P < 0,01$; *** $P < 0,001$).

In generale si osserva un forte LD in tutti i campioni tra la prima posizione polimorfica localizzata all'estremità 5'UTR (rs779804) e l'ultima posizione polimorfica in regione 3'UTR (rs1642741). La stessa forte associazione si riscontra tra i tre siti polimorfici rs1642742, rs1681669 e rs1681668. Poiché dunque possiamo considerare l'rs779805 e l'rs1642742 come rappresentativi (tagSNP) degli SNP con cui sono in LD, in caso di test di associazione abbiamo utilizzato solo 3 dei 6 SNP totali, come spiegato nei paragrafi successivi. Sia le popolazioni di alta quota che le popolazioni di bassa quota sono in equilibrio di Hardy-Weinberg.

In tabella 16 sono riportate le frequenze degli aplotipi stimati con un metodo di ricostruzione della fase e in tabella 17 le stime degli indici di variabilità.

<u>VHL</u>	CS+CB+PN	Cusco (CS)	Cusibamba (CB)	Puno (PN)	Maya
N. di cromosomi	178	56	32	90	50
N. di siti	1891	1891	1891	1891	1891
N. di siti polimorfici	6	6	6	6	6
N. di aplotipi	6	4	5	3	4
Diversità aplotipica $H \pm SD$	$0,67 \pm 0,01$	$0,66 \pm 0,04$	$0,73 \pm 0,03$	$0,65 \pm 0,02$	$0,55 \pm 0,06$
$\pi \pm SD (x10^{-3})$	$1,50 \pm 0,02$	$1,53 \pm 0,06$	$1,41 \pm 0,13$	$1,51 \pm 0,04$	$1,22 \pm 0,13$
$\theta_s \pm SD$ (per sites)	$1,04 \pm 0,47$	$1,31 \pm 0,62$	$1,49 \pm 0,73$	$1,18 \pm 0,55$	$1,34 \pm 0,64$
Tajima's D	3,56*	3,02*	2,24*	3,23*	1,85
Fu&Li's D	1,08	1,19	1,25	1,14	1,20

Tab.17 *VHL*: Indici di variabilità genetica intra-popolazione.

* p-value<0,05

Analizzando i valori di questi indici nei tre siti di campionamento si può notare che il sito di Cusibamba mostra valori di alcuni indici di diversità come H e θ_s maggiori rispetto alle popolazioni di Puno e Cusco. Questo risultato è in contrasto con i risultati degli stessi test effettuati per sequenze del mtDNA, per le quali la popolazione di Cusibamba mostrava valori più bassi rispetto alle altre popolazioni, come atteso nel caso di isolamento geografico.

Rispetto alla HVRI del mtDNA sono evidenti alcune differenze anche per le popolazioni di Cusco e Puno, in particolare sono da evidenziare i risultati positivi dei test di neutralità (D

di Tajima (Tajima 1989) per il quale i risultati sono significativo, e *D di Fu e Li* (Fu and Li 1993)). Questo tipo di risultato è indice di un eccesso di varianti genetiche mantenute a frequenza intermedia nelle popolazioni, potrebbe essere interpretato come l'effetto dell'azione della selezione bilanciante o di particolari fenomeni demografici come colli di bottiglia o struttura interna delle popolazioni (Nielsen 2001).

8.2.2 VHL: confronti alta e bassa quota

Gli indici di diversità aplo-tipica e nucleotidica indicano livelli di variabilità genetica interna maggiori negli individui andini rispetto ai Maya. I risultati dei test di neutralità mostrano valori positivi per entrambe le popolazioni, tuttavia negli andini il *D di Tajima* raggiunge valori significativi tra 2,24 e 3,56 rispetto ad un valore di 1,85 non significativo riscontrato nella popolazione di bassa quota.

La misura delle distanze genetiche effettuata calcolando l'indice F_{ST} a coppie tra le popolazioni del campione di alta quota (Andini) e il campione di bassa quota (Maya) riporta un valore significativo pari al 6% indicando differenziazione genetica per gli SNP considerati (Tab.18)

	Cusco	Cusibamba	Puno	Maya
Cusco	0			
Cusibamba	0,022	0		
Puno	0	0,008	0	
Maya	0,036	0,148*	0,052*	0
Andini ⁺	-	-	-	0,062*

Tab.18 VHL: F_{ST} a coppie

* p-value < 0,05

⁺ Andini: si intendono le popolazioni di Cusco, Cusibamba e Puno considerate tutte insieme.

Nello specifico se vengono confrontate le singole popolazioni del campione di alta quota con il campione di bassa quota, si osservano valori di F_{ST} significativi nei confronti Maya-Puno e Maya-Cusibamba e in particolare quest'ultimo riporta un valore di F_{ST} pari al 14,8%, mentre i tre campioni di alta quota sono fra di loro molto omogenei (Tab.18).

Ciò indica una tendenza alla divergenza tra le popolazioni di alta quota e il campione di bassa quota per quanto riguarda le varianti genetiche identificate dal nostro studio al locus *VHL*.

Per confermare questa ipotesi sono stati effettuati dei test di esatti di Fisher allo scopo di verificare se esista associazione tra un determinato genotipo e l'appartenenza al gruppo di alta quota o di bassa quota. Per via del totale LD tra alcuni SNP, come spiegato precedentemente, le analisi sono state effettuate solo per 3 siti polimorfici le cui frequenze sono rappresentative per gli altri siti.

Nello specifico gli SNP testati sono *rs779804*, *rs779805* e *rs1642742*.

I p-value dei test effettuati sono compresi tra 0,005 e 0,03 confermando che ci sono delle differenze significative nella frequenze genotipiche tra i campioni di alta quota e di bassa quota. Osservando le tabelle di contingenza si può notare come tale associazione possa essere dovuta ad un eccesso di individui eterozigoti nel campione di alta quota rispetto al campione di bassa quota. Un aumento del numero di individui eterozigoti potrebbe essere un effetto imputabile alla selezione bilanciante del tipo risultante da un vantaggio dell'eterozigote, anche se l'esistenza di equilibrio di Hardy-Weinberg non permette di considerare definitiva questa spiegazione.

8.2.3 *VHL*: associazione genotipo/fenotipo

Alla luce di questi risultati sono state effettuate le analisi di associazione genotipo/fenotipo all'interno del campione Andino. Data la bassa distanza genetica tra i 3 campioni di alta quota (tabella 19) e i risultati di analisi preliminari hanno permesso di escludere evidenze di stratificazione del campione "Ande" in base all'età e all'etnia, i test di associazione sono stati effettuati considerando tutti gli individui come un unico campione.

Come nel caso di *EPO* sono stati effettuati due tipi di analisi:

- l'analisi della varianza (ANOVA) per testare l'associazione genotipo/fenotipo considerando gli individui del campione suddivisi in base al sesso (55maschi, 34 femmine).

- il test esatto di Fisher per confrontare i genotipi appartenenti a categorie fenotipiche estreme determinate dalla suddivisione in quartili della distribuzione ordinale dei valori della variabile fenotipica quantitativa. Queste analisi sono state effettuate solo per il gruppo delle femmine poiché l'esigua numerosità nel campione maschile determina la presenza di un numero elevato di celle con valore 0 a seguito della suddivisione in quartili e in categorie genotipiche.

Le variabili quantitative continue prese in esame sono le stesse analizzate anche per il gene *EPO*: indice di massa corporea (BMI), pressione sistolica (Sis) e pressione diastolica (Dia), frequenza del battito cardiaco (BPM), percentuale di saturazione di ossigeno (SpO₂), temperatura corporea (T°), emoglobina (Hb), percentuale di ematocrito (HTC), livelli di glucosio misurati solo per il campione di Puno.

Nessuno dei test effettuati ha riportato valori statistici significativi. Questo significa che analizzando il campione di alta quota al suo interno non è possibile rilevare associazioni tra i genotipi e le variabili fenotipiche misurate.

8.3 *EPAS1 (HIF-2a)*

Il sequenziamento del gene *EPAS1* è stato limitato alle regioni dove sono localizzati 5 polimorfismi (*rs7583554*, *rs7583088*, *rs11678465*, *rs7557402* e NCBI Build 36.1: posizione 46441523 C>G) che mostrano associazione con i livelli di ematocrito. 4 polimorfismi sono stati individuati da Beall et al. (2010) e 1 polimorfismo individuato da Yi et al. (2010). Tali polimorfismi sono tutti intronici pertanto abbiamo sequenziato 1800 bp del gene suddivise in due frammenti di 972bp e 828bp comprendenti l'introne 4, l'introne 5 e l'introne 9 in 16 individui (32 cromosomi) del campione di Cusibamba.

Tutti i campioni tipizzati sono monomorfici e in particolare non presentano le varianti polimorfiche associate a bassi livelli di emoglobina. Questo risultato supporta ulteriormente l'ipotesi di due percorsi evolutivi differenti seguiti da popolazioni andine e tibetane per raggiungere il medesimo scopo: vivere in condizioni di basse pressioni di ossigeno.

9. DISCUSSIONE E CONCLUSIONI

Il duplice scopo degli studi di genetica di popolazione applicati all'uomo è quello di spiegare i modelli di diversità genetica in termini di storia delle popolazioni, e al tempo stesso di capire le basi genetiche della variabilità fenotipica che ha permesso agli esseri umani di colonizzare ambienti tanto diversi. Capire come la variabilità genetica influisce sulla capacità dell'uomo di adattarsi a diversi ambienti, o regioni geografiche, ha importanti risvolti sia in campo evolutivo, sia in campo medico. Ne sono una dimostrazione gli studi che hanno portato alla comprensione dei meccanismi evolutivi alla base di fenotipi quali la tolleranza al lattosio (Enattah et al. 2002, Lewinsky et al. 2005, Tishkoff et al. 2007), la resistenza alla malaria (Tishkoff et al. 2001, Saunders et al. 2002, Louicharoen et al. 2009), la regolazione della pigmentazione (Sturm 1998, Sturm et al. 1998, Harding et al. 2000) o la percezione del gusto amaro (Wooding et al. 2004, Kim et al. 2005, Soranzo et al. 2005).

In questo studio abbiamo concentrato la nostra attenzione sulla comprensione delle basi genetiche legate all'adattamento dell'uomo alla vita in alta quota dove la bassa pressione di ossigeno determina uno stato di ipossia che esercita sugli organismi un forte stress fisiologico.

I primi studi su popolazioni native di alta quota risalgono già alla metà del secolo scorso (Monge 1942, Matthews 1945, Valdivia 1961, Baker 1969, Morpurgo et al. 1976). Questi studi portarono ad indentificare una serie di differenze fenotipiche sia tra nativi di alta quota e nativi di bassa quota sia tra popolazioni native di alta quota provenienti da diverse aree geografiche, Andini e Tibetani. Alcuni di questi fenotipi, come l'elevata concentrazione di emoglobina in popolazioni andine o l'aumentata risposta ventilatoria in popolazioni tibetane, sono stati considerati adattativi poiché hanno evidenziato un elevato livello di ereditabilità in queste popolazioni. Per questo gli studi successivi si sono concentrati sullo studio della variabilità genetica alla base di questi fenotipi.

Negli ultimi anni numerosi studi di geni candidati sono stati effettuati senza tuttavia portare a conclusioni definitive (Rupert et al. 2003, Bigham et al. 2008, Stobdan et al. 2008).

Questo ha spinto il nostro gruppo di ricerca ad avviare un nuovo progetto studiando popolazioni native delle Ande peruviane. Il nostro gruppo di ricerca si occupa ormai da anni dello studio della variabilità genetica di popolazioni andine (Tarazona-Santos et al. 2001, Fuselli et al. 2003, Fuselli et al. 2007) in collaborazione con il gruppo di Eduardo Tarazona-Santos dell'università di Belo Horizonte in Brasile. Questo progetto pertanto ha permesso di ampliare le conoscenze legate a queste popolazioni in relazione anche ad aspetti riguardanti l'adattamento all'ambiente di alta quota.

Rispetto ai precedenti studi, in questo progetto abbiamo utilizzato un metodo innovativo per la scelta dei geni candidati. Attraverso strumenti bioinformatici capaci di integrare le informazioni provenienti dalla letteratura con quelle dei database pubblici è stato possibile individuare non solo geni direttamente coinvolti con un fenotipo, ma anche geni appartenenti a cicli metabolici che interagendo tra loro possono determinare la manifestazione di specifici fenotipi legati all'adattamento all'ipossia. Questo metodo si è dimostrato efficace anche in altri studi che coinvolgono associazioni tra geni e fenotipi complessi, come l'adattamento al clima (Hancock et al. 2008).

Mediante questo approccio è stata identificata una lista di geni candidati che fanno capo al ciclo metabolico degli Hypoxia Inducible Factors (HIF).

La validità dei nostri risultati nella scelta dei geni candidati è stata confermata anche da studi pubblicati recentemente in letteratura ed effettuati con approcci genomici (Beall et al. 2010, Bigham et al. 2010, Simonson et al. 2010, Yi et al. 2010). In questi studi infatti i polimorfismi che mostrano segnali selettivi sono concentrati in regioni introniche o intergeniche di geni, quali *EPAS1* e *EGLN1*, che appartengono alla via metabolica di HIF e rientrano nella lista di geni candidati da noi creata.

Inoltre allo scopo di indirizzare lo studio verso le regioni dei geni la cui variabilità fosse più probabilmente associata a fenotipi di adattamento all'ipossia abbiamo effettuato uno studio di genetica comparativa analizzando la variabilità di questi loci a livello interspecifico. Questo ha permesso sia di comprendere sotto quali pressioni selettive si sono evolute le regioni codificanti di questi geni sia di verificare la presenza di segnali di selezione diversificante in specie adattate ad ambienti ipossici e specie che vivono in condizioni di normossia.

Le analisi sono state effettuate utilizzando il software PAML che rappresenta oggi lo strumento migliore per rilevare l'azione della selezione positiva in analisi comparative, basate sulla stima dei tassi di sostituzione sinonima e non-sinonima e del loro rapporto definito ω (Nielsen 2005).

Le analisi effettuate per verificare come varia ω lungo la sequenza codificante dei geni candidati di alcuni mammiferi, hanno evidenziato un generale livello di conservazione di tali sequenze con molti siti sottoposti all'azione della selezione purificante volta a mantenere ridotta la variabilità non sinonima e pertanto inalterata la struttura e la funzione della proteina codificata. Questo conferma che si tratta di geni che svolgono ruoli chiave nei processi fisiologici ai quali appartengono per cui sia necessario il mantenimento di un'integrità della struttura proteica, in specie di mammiferi che hanno in comune lo stesso sistema respiratorio. Studi recenti hanno infatti dimostrato l'esistenza di diversi tassi evolutivi per geni quali HIF-1 α e HIF-2 α , quando si confrontano specie con sistemi respiratori estremamente differenti come teleostei e mammiferi indicando per i primi un tasso evolutivo maggiore rispetto a quello riscontrato nei mammiferi (Rytönen et al. 2011), associato appunto al sistema respiratorio in ambiente acquatico.

Ambienti come l'alta quota o l'ambiente sotterraneo sono comunque caratterizzati dalla presenza di basse pressioni di ossigeno che creano una condizione di stress alla quale non tutte le specie di mammifero sono state in grado di adattarsi. Nel nostro studio quindi abbiamo cercato di verificare se questo ciclo metabolico così conservato all'interno del gruppo dei mammiferi mostrasse tuttavia tassi di evoluzione differenti in specie adattate a vivere in ambienti a basse pressioni di ossigeno e specie che vivono in condizioni di normossia.

Il confronto tra specie che vivono in presenza di livelli normali di ossigeno e specie adattate all'ipossia raggruppate in un unico dataset non ha rilevato differenze nei tassi di evoluzione per i geni EPO ed HIF-1 α . Tuttavia analizzando ciascuna specie adattata all'ipossia indipendentemente dalle altre sono stati rilevati segnali di selezione diversificante tra il gruppo delle specie che vivono in normossia e il *Bos grunniens* (Yak) per il gene HIF-1 α .

Il *Bos grunniens*, comunemente noto come Yak, o bue tibetano, è una specie domestica appartenente alla famiglia dei bovidi e diffusa nell'altopiano tibetano a quote comprese tra i 3500 m e i 5000 m. La rispettiva specie selvatica è il *Bos mutus*. Secondo alcuni studi il *Bos mutus* e le specie *Bison bison* (Bisonte americano) del nord America si sono evoluti in Asia centrale a da un antenato comune e la divergenza tra le due specie risale a circa 2 milioni di anni fa. Secondo questa teoria lo Yak è rimasto in Asia centrale principalmente nelle regioni degli altopiani, mentre le specie del genere *Bison* si sono diffuse in Nord America attraverso la regione della Beringia, tra la metà e la fine del Pleistocene (126 Kya -11,7 Kya) (Leslie and Schaller 2009). Rispetto alle altre specie del genere *Bos*, il *Bos grunniens* mostra evidenti caratteristiche di adattamento all'alta quota come una maggiore eritrocitosi, una superficie alveolare maggiore, e ridotta vasocostrizione ipossica (Anand et al. 1986, Williams 1994). Non sono note chiare associazioni tra questi fenotipi e specifici loci, tuttavia il risultato di questa analisi potrebbe indicare un ruolo importante svolto appunto da HIF-1 α . Il risultato ottenuto con le nostre analisi infatti evidenzia un'accelerazione significativa dell'evoluzione nel ramo di questa specie adattata all'ipossia rispetto alle altre specie che vivono in condizioni di normossia.

Il risultato non è stato ottenuto per le altre specie analizzate adattate all'ipossia, tuttavia questo potrebbe dipendere dal fatto che si tratta di specie che vivono in ambienti estremamente differenti come l'ambiente acquatico per il *Tursiops truncatus* e l'ambiente sotterraneo per le il genere *Spalax*. Nell'analisi è stata inclusa anche la specie *Vicugna pacos* (*Alpaca*), anch'essa una specie domestica nativa di alta quota, diffusa negli altopiani andini e appartenente all'ordine dei camelidi. I camelidi sudamericani si sono originati in Nord America circa 40 milioni di anni fa e gli antenati della vigogna e del guanaco sono migrati in Sud America intorno a 2 milioni di anni fa (Wheeler et al. 2006). L'alpaca rappresenta la specie domestica della vigogna (Mengoni Goñalons 2008).

La mancanza di un segnale significativo potrebbe dipendere dal fatto che si tratta di specie filogeneticamente distanti che hanno colonizzato gli ambienti di alta quota da tempi differenti e pertanto si può escludere l'ipotesi di un adattamento convergente.

Tuttavia in futuro sarà necessario replicare le analisi per verificare la stabilità del risultato poiché uno dei maggiori problemi riscontrati in questo tipo di studi è rappresentato dall'affidabilità delle sequenze depositate nei database pubblici. In questo studio abbiamo utilizzato una serie di restrizioni nella scelta delle sequenze allo scopo di ridurre al minimo i falsi positivi. Seguendo questa logica, la sequenza di HIF-1 α del *Bos grunniens* utilizzata in questo studio è quella pubblicata nel lavoro di Wang et al. (2006) la cui sequenza amminoacidica è nel database UniProt (<http://www.uniprot.org/uniprot/Q0PPG7>). Tuttavia nel lavoro di Dolt et al. (2007) è stata pubblicata una sequenza che mostra un numero inferiore di sostituzioni non sinonime rispetto alla sequenza della specie più vicina, il *Bos taurus*, e pertanto potrebbe determinare una stima inferiore del tasso di evoluzione del *Bos grunniens* rispetto a quella stimata in questa prima fase dello studio. Il motivo della presenza in letteratura di due sequenze così differenti non è chiaro, ma potrebbe essere il risultato della provenienza geografica degli individui analizzati nei due lavori, gli uni provenienti dall'altopiano dell'altopiano del Qingha in Tibet (Wang et al. 2006), l'altro dalla regione indiana di Ladakh, Jammu and Kashmir (Dolt et al. 2007) distante dalla prima più di mille Km.

I risultati delle analisi interspecifiche, che hanno dimostrato che le regioni codificanti dei geni candidati sono generalmente molto conservate, sono stati confermati dalle analisi delle regioni codificanti dei geni candidati *EPO* e *VHL*, dai quali siamo partiti per iniziare lo studio della variabilità genetica in popolazioni umane di alta quota.

Questo ha permesso di focalizzare l'attenzione sulle regioni regolative di questi geni e in particolare sulle regioni 5' e 3' UTR.

Per quanto riguarda il gene *EPO* i due polimorfismi riscontrati nelle regioni regolative a monte e a valle del gene non mostrano differenze significative nelle frequenze alleliche né all'interno del campione di alta quota né confrontando il campione di alta quota con il rispettivo controllo di bassa quota. I test di associazione genotipo/fenotipo non hanno mostrato associazioni significative tra questi polimorfismi e le variabili fenotipiche misurate per i campioni di alta quota.

In particolare, diversamente da quanto riscontrato nel lavoro di Tong et al. (2008) non si rileva alcuna associazione tra la variante A di rs1617640 ed un incremento dell'espressione del gene rappresentata da elevati livelli di ematocrito. Infatti sia i test di associazione effettuati considerando i tutti gli individui di alta quota sia le analisi dei fenotipi estremi non hanno riportato valori significativi indicando che i valori di ematocrito sono indipendenti dal genotipo alla variante rs1617640 all'interno del campione di alta quota.

Inoltre, se come atteso secondo il lavoro di Tong et al. (2008), l'allele A fosse associato a maggiori livelli di ematocrito, ci aspetteremmo di riscontrare differenze significative nelle frequenze alleliche tra popolazioni di bassa quota e popolazioni di alta quota dal momento che queste ultime presentano valori di ematocrito significativamente maggiori di quelli presenti in letteratura per popolazioni di bassa quota. Come già detto, i campioni Maya e i campioni andini tipizzati in questo studio mostrano le stesse frequenze alleliche. Quindi, pur non avendo a disposizione le misurazioni fenotipiche di ematocrito nei campioni di bassa quota tipizzati (Maya) è possibile ipotizzare che i polimorfismi individuati non influenzino in modo significativo i livelli di ematocrito.

Studi recenti hanno dimostrato l'associazione di polimorfismi localizzati nelle regioni non codificanti del gene *EPAS1* (HIF-1 α) e i livelli di ematocrito in popolazioni tibetane (Beall et al. 2010, Yi et al. 2010). Pertanto è possibile che la modulazione dei livelli di ematocrito non dipenda principalmente da polimorfismi presenti nelle regioni regolative del gene *EPO*, ma piuttosto da una maggiore o minore attività del fattore di trascrizione che ne controlla l'espressione.

Nei lavori sopra citati, trattandosi di popolazioni tibetane, i polimorfismi del gene *EPAS1* che hanno mostrato segnali significativi sono associati al mantenimento di bassi livelli emoglobina, e quindi di ematocrito. In accordo con le teorie che sostengono un'evoluzione parallela di andini e tibetani (Beall 2007b), tali polimorfismi sono assenti nei nostri campioni i quali mostrano valori elevati di ematocrito.

Allo stesso modo è possibile che esistano polimorfismi in alcune regioni del gene *EPAS1* associabili agli elevati livelli di ematocrito in popolazioni andine, e che tali polimorfismi non siano stati riscontrati negli studi effettuati su popolazioni tibetane. Per questo motivo in futuro saranno sequenziate altre regioni del gene *EPAS1* nei nostri campioni andini.

Pertanto considerando sia le analisi di variabilità interna alle popolazioni, sia tra popolazioni e i risultati delle analisi di associazione possiamo concludere che i polimorfismi individuati nel gene *EPO* nei campioni tipizzati in questo studio non sono associabili all'adattamento alle basse pressioni di ossigeno.

Risultati incoraggianti invece sono stati riscontrati analizzando il gene *VHL*.

Le regioni 5' e 3' UTR del gene mostrano la presenza di 6 siti polimorfici in elevato LD tra loro. Questi polimorfismi presentano frequenze differenti nei campioni di alta quota e bassa quota e in particolare i valori di diversità nucleotidica (π) stimati per le popolazioni andine sono maggiori rispetto a quelli delle popolazioni Maya di bassa quota.

Valori elevati di diversità nucleotidica sono generalmente attesi in regioni influenzate dall'azione della selezione bilanciante. In caso di selezione bilanciante si osserva un eccesso di varianti mantenute nelle popolazioni a frequenze intermedie generalmente dovuto ad un vantaggio del genotipo eterozigote.

Pertanto se effettivamente i valori elevati di π nelle popolazioni andine fossero dovuti alla selezione bilanciante ci aspetteremmo di osservare delle differenze nelle frequenze genotipiche tra campioni di alta quota e campioni di bassa quota. I risultati dei test di Fisher sono in accordo con questa ipotesi indicando differenze significative ($0,005 < p\text{-value} < 0,03$) tra i due campioni e in particolare si osserva un eccesso di individui eterozigoti nel campione di alta quota rispetto al campione di bassa quota.

Anche il valore F_{ST} utilizzato come misura di distanze genetiche tra popolazioni, indica che le popolazioni di alta quota si differenziano significativamente dalle popolazioni di bassa quota. Questo risultato potrebbe dipendere dalle differenze tra popolazioni dovute alla demografia, tuttavia le stime di F_{ST} effettuate tra popolazioni Maya e andine per i marcatori neutrali (microsatelliti) tipizzati nel lavoro di Wang et al. (2007) indicano un basso livello di differenziazione tra le due popolazioni. E' importante però ricordare che il confronto diretto tra questi due tipi di marcatori potrebbe essere influenzato dal fatto che i microsatelliti sono marcatori con un modello mutazionale diverso dalle sequenze nucleari, in particolare sono caratterizzati da tassi di mutazione decisamente più elevati. In ogni caso per confermare i risultati ottenuti in futuro queste regioni del gene *VHL*

saranno tipizzate anche in un altro campione di bassa quota composto di 25 individui provenienti dalle popolazioni amazzoniche di Andoas di Loreto e Lamas di San Martin, più vicine geograficamente alle popolazioni andine.

L'ipotesi di selezione bilanciante è supportata anche dai risultati ottenuti con i test di neutralità D di Tajima (Tajima 1989) e D di Fu&Li (Fu and Li 1993).

I valori dei test di neutralità sono positivi sia per le popolazioni di alta quota, sia per le popolazioni di bassa quota, tuttavia i test D di Tajima sono significativi con $p\text{-value} < 0,02$ solo nelle popolazioni di alta quota. In particolare questi test sono significativi anche per il campione di Cusibamba, che presenta una dimensione campionaria più piccola (16 individui) rispetto al campione Maya (25 individui), escludendo dunque l'ipotesi che la dimensione dei campioni determini una differenza tra quote.

Come discusso nel capitolo riguardante le analisi statistiche (paragrafo 4.2), risultati di questo tipo possono essere dovuti non solo ad eventi di selezione bilanciante, ma anche a fenomeni demografici quali riduzioni delle dimensioni delle popolazioni (collo di bottiglia) o suddivisione delle popolazioni stesse (Przeworski et al. 2000, Nielsen 2001). Per distinguere tra le diverse ipotesi è necessario verificare quali eventi demografici hanno caratterizzato le popolazioni in esame, idealmente tipizzando numerose regioni neutrali del genoma. Questo perché la selezione agisce su loci singoli, mentre la demografia coinvolge tutto il genoma, quindi analizzando più loci ci si aspetta che, esclusi alcuni loci selezionati, gli altri diano un'idea della storia della popolazione (Nielsen 2001, Nielsen et al. 2007).

Pur non avendo a disposizione informazioni relative a regioni neutrali del genoma nucleare, possiamo ipotizzare che i valori positivi riscontrati nei test di neutralità siano imputabili all'azione della selezione bilanciante confrontandoli con i risultati ottenuti per la regione neutrale HVRI del mtDNA. I risultati di queste analisi, in accordo con studi precedenti (Fuselli et al. 2003), riportano valori negativi significativi compatibili con espansione demografica delle popolazioni. Sulla base di questo quindi si potrebbe ipotizzare che il risultato ottenuto per *VHL* nelle stesse popolazioni sia effettivamente imputabile all'azione della selezione bilanciante. Tuttavia si deve tener conto del fatto che

il mtDNA e il DNA nucleare hanno dimensioni effettive e tassi di mutazione diversi, dunque il paragone è plausibile, ma è anche possibile che, dati questi due fattori confondenti, il confronto diretto possa essere fuorviante. Pertanto sono in prospettiva analisi per verificare se le differenze osservate tra Andini e Maya al locus *VHL* sono attese o meno in condizioni di neutralità attraverso studi di simulazione basati sul coalescente. Se la nostra ipotesi venisse confermata potremmo ipotizzare che mantenere diversi alleli a frequenza intermedia sia in qualche modo vantaggioso per la vita in alta quota.

I test di associazione genotipo/fenotipo effettuati all'interno del campione di alta quota non hanno mostrato risultati significativi. Seppure al momento non siano disponibili misurazioni fenotipiche per il campione di bassa quota, tali da poter effettuare i test di associazione, le differenze genotipiche riscontrate tra individui di alta e individui di bassa inducono ad ulteriori speculazioni.

In che termini potrebbe essere vantaggioso mantenere un bilanciamento delle frequenze alleliche a questo locus in popolazioni di alta quota?

Al momento non sono note associazioni di questi polimorfismi con particolari alterazioni della struttura proteica.

Come è stato descritto in questa tesi la corretta regolazione del ciclo di HIF ha un ruolo fondamentale nell'attivazione di una risposta fisiologica all'ipossia. Tuttavia se lo stato di ipossia permane per tempi troppo lunghi, come nel caso di patologie tumorali, la risposta a tale stress mediata da HIF contribuisce ad aumentare l'attività cellulare e può favorire la proliferazione tumorale. Similmente in individui sani sottoposti ad ipossia per lunghi periodi la sovra-regolazione del ciclo di HIF determinerebbe l'attivazione costitutiva di una via metabolica troppo dispendiosa per il bilancio energetico della cellula.

In alta quota la condizione di ipossia è cronica, pertanto devono esistere dei meccanismi che bilanciano la concentrazione di HIF- α all'interno della cellula senza ridurla come in condizioni di normossia, ma senza neanche aumentarla come accade nelle cellule tumorali.

In condizioni di normossia i livelli di HIF- α vengono mantenuti normali grazie al processo di degradazione mediato da VHL. Tale processo è reso possibile dalla presenza di ossigeno che permette la reazione di idrossilazione dei due residui di prolina presenti nei domini ODD e N-TAD di HIF- α . Se tali domini non sono idrossilati, come in condizioni di ipossia, VHL non è in grado di legare HIF- α .

Ad esempio Sufan et al. (2009), allo scopo di ripristinare un sistema di soppressione tumorale, hanno ingegnerizzato una sequenza di VHL in grado di legare HIF- α anche in condizioni di ipossia cronica favorendone la degradazione e la conseguente riduzione dell'attività proliferativa tumorale.

Ciò non è tuttavia compatibile con quanto da noi riscontrato, poiché i polimorfismi studiati sono localizzati in regioni che non controllano il legame di VHL ad HIF- α .

Se non si tratta di cambiamenti proteici si potrebbe ipotizzare un controllo in cis dell'espressione genica. Numerosi studi hanno infatti evidenziato che i fattori genetici che contribuiscono maggiormente alla variabilità dei tratti complessi risiedono proprio nelle regioni di regolazione, piuttosto che nelle porzioni codificanti del genoma (Macdonald and Long 2005), e molto spesso sono soggette all'azione della selezione bilanciante (Bamshad et al. 2002, Fagundes et al. 2005, Wilson et al. 2006, Donadi et al. 2011)

Il controllo post-trascrizionale degli mRNA, dallo splicing alla traduzione, rappresenta uno dei punti fondamentali nella modulazione del profilo di espressione proteica (Koritzinsky et al. 2006, van den Beucken et al. 2006) e permette alla cellula di reagire a stress ambientali, quali le variazioni nella concentrazione di ossigeno (Sonenberg and Hinnebusch 2009), in modo rapido ed efficace. Tale controllo è mediato da RNA Binding Protein (RBP) capaci di legare specifiche sequenze presenti nelle sequenze di mRNA.

Supponiamo quindi che i cambiamenti in 5' e 3' UTR permettano un legame VHL-RBP che ne diminuisca la traduzione. Individui eterozigoti avrebbero una quantità intermedia di proteina rispetto ai due omozigoti "wt-wt" e "mut-mut" (LaJevic et al. 2010, Jafarifar et al. 2011).

In condizioni di estrema ipossia (tumori) HIF- α non viene idrossilato, non viene quindi degradato da VHL e la via metabolica a valle è del tutto attivata.

In condizioni di ipossia moderata (alta quota) assumiamo che una certa concentrazione di molecole di HIF- α venga idrossilata, dunque degradata via VHL, cosicché la via metabolica a valle venga attivata in modo meno pronunciato rispetto ad una condizione patologica. In questo caso una modulazione "fine" dell'espressione di VHL ricadrebbe sul controllo "fine" dell'attività di HIF. Queste condizioni sono tipicamente quelle in cui un controllo di tipo regolativo permette una migliore modulazione del fenotipo rispetto ad esempio, ad un cambiamento amminoacidico.

Inoltre, indipendentemente dal ruolo di controllo della degradazione di HIF, il coinvolgimento di VHL in numerose vie di regolazione cellulari (Galban et al. 2003, Stickle et al. 2005, Danilin et al. 2009, Danilin et al. 2010) in condizioni tumorali mostra come una modulazione della sua espressione possa conferire un livello ulteriore di regolazione dell'attività della cellula in condizioni di bassa pressione di ossigeno.

Studi futuri si concentreranno quindi sull'analisi di queste regioni allo scopo di comprendere se i siti polimorfici individuati nei nostri campioni risiedano appunto in specifiche sequenze di riconoscimento per RBP e possano quindi essere associati ad una migliore regolazione dell'espressione genica in condizioni di basse pressioni di ossigeno. Questo sarà verificato anche effettuando studi di espressione in vitro.

Inoltre in futuro si procederà al sequenziamento e all'analisi degli altri geni candidati individuati in questo studio, a partire da *EGLN1* e da regioni di *EPAS1* non ancora investigate nei nostri campioni, che hanno evidenziato segnali di associazione con fenotipi di adattamento all'ipossia in altri studi.

Lo studio delle basi genetiche di un processo di adattamento complesso come quello descritto in questa tesi rappresenta quindi il punto di partenza per una maggiore comprensione dei processi alla base dell'evoluzione dell'uomo e delle patologie dovute ad un mancato adattamento all'ambiente o all'alterazione dei sistemi cellulari di risposta a condizioni di stress.

Riferimenti bibliografici

- Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR *et al.* The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 2008; 3: e1764.
- Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F *et al.* Gene prioritization through genomic data fusion. *Nat Biotechnol* 2006; 24: 537-544.
- Akey JM. Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res* 2009; 19: 711-722.
- Anand IS, Harris E, Ferrari R, Pearce P, Harris P. Pulmonary haemodynamics of the yak, cattle, and cross breeds at high altitude. *Thorax* 1986; 41: 696-700.
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J *et al.* Sequence and organization of the human mitochondrial genome. *Nature* 1981; 290: 457-465.
- Appenzeller O, Minko T, Qualls C, Pozharov V, Gamboa J, Gamboa A *et al.* Gene expression, autonomic function and chronic hypoxia: lessons from the Andes. *Clin Auton Res* 2006; 16: 217-222.
- Baker PT. Human adaptation to high altitude. *Science* 1969; 163: 1149-1156.
- Baker PT. The Adaptive Limits of Human Populations. *Man, New Series* 1984; 19: 1-14.
- Balaresque PL, Ballereau SJ, Jobling MA. Challenges in human genetic diversity: demographic history and adaptation. *Hum Mol Genet* 2007; 16 Spec No. 2: R134-139.
- Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS *et al.* A strong signature of balancing selection in the 5' cis-regulatory region of CCR5. *Proc Natl Acad Sci U S A* 2002; 99: 10539-10544.
- Banks RE, Tirukonda P, Taylor C, Hornigold N, Astuti D, Cohen D *et al.* Genetic and epigenetic analysis of von Hippel-Lindau (VHL) gene alterations and relationship with clinical variables in sporadic renal cancer. *Cancer Res* 2006; 66: 2000-2011.
- Beall CM. Tibetan and Andean patterns of adaptation to high-altitude hypoxia. *Hum Biol* 2000; 72: 201-228.
- Beall CM. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integrative and Comparative Biology* 2006; 46: 18-24.
- Beall CM. Detecting natural selection in high-altitude human populations. *Respir Physiol Neurobiol* 2007a; 158: 161-171.
- Beall CM. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc Natl Acad Sci U S A* 2007b; 104 Suppl 1: 8655-8660.
- Beall CM, Brittenham GM, Strohl KP, Blangero J, Williams-Blangero S, Goldstein MC *et al.* Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *Am J Phys Anthropol* 1998; 106: 385-400.
- Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J *et al.* Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci U S A* 2010; 107: 11459-11464.
- Beall CM, Strohl KP, Blangero J, Williams-Blangero S, Almasy LA, Decker MJ *et al.* Ventilation and hypoxic ventilatory response of Tibetan and Aymara high altitude natives. *Am J Phys Anthropol* 1997; 104: 427-447.
- Bellance N, Lestienne P, Rossignol R. Mitochondria: from bioenergetics to the metabolic regulation of carcinogenesis. *Front Biosci* 2009; 14: 4015-4034.
- Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet* 2004; 74: 1111-1120.
- Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R *et al.* Identifying signatures of natural selection in tibetan and andean populations using dense genome scan data. *PLoS Genet* 2010; 6:

- Bigham AW, Kiyamu M, Leon-Velarde F, Parra EJ, Rivera-Ch M, Shriver MD *et al.* Angiotensin-converting enzyme genotype and arterial oxygen saturation at high altitude in Peruvian Quechua. *High Alt Med Biol* 2008; 9: 167-178.
- Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG *et al.* Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. *Hum Genomics* 2009; 4: 79-90.
- Bonato SL, Salzano FM. Diversity and age of the four major mtDNA haplogroups, and their implications for the peopling of the New World. *Am J Hum Genet* 1997; 61: 1413-1423.
- Breen E, Tang K, Olfert M, Knapp A, Wagner P. Skeletal muscle capillarity during hypoxia: VEGF and its activation. *High Alt Med Biol* 2008; 9: 158-166.
- Brown DE, Beall CM, Strohl KP, Mills PS. Exhaled nitric oxide decreases upon acute exposure to high-altitude hypoxia. *Am J Hum Biol* 2006; 18: 196-202.
- Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L *et al.* A human genome diversity cell line panel. *Science* 2002; 296: 261-262.
- Cann RL, Wilson AC. Length mutations in human mitochondrial DNA. *Genetics* 1983; 104: 699-711.
- Castillo O, Woolcott OO, Gonzales E, Tello V, Tello L, Villarreal C *et al.* Residents at high altitude show a lower glucose profile than sea-level residents throughout 12-hour blood continuous monitoring. *High Alt Med Biol* 2007; 8: 307-311.
- Castrop H, Kurtz A. Functional evidence confirmed by histological localization: overlapping expression of erythropoietin and HIF-2alpha in interstitial fibroblasts of the renal cortex. *Kidney Int* 2010; 77: 269-271.
- Clark PE, Cookson MS. The von Hippel-Lindau gene: turning discovery into therapy. *Cancer* 2008; 113: 1768-1778.
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B *et al.* The Last Glacial Maximum. *Science* 2009; 325: 710-714.
- Cogo A. *Medicina e salute in montagna*. Hoepli; 2009.
- Danilin S, Sourbier C, Thomas L, Lindner V, Rothhut S, Dormoy V *et al.* Role of the RNA-binding protein HuR in human renal cell carcinoma. *Carcinogenesis* 2010; 31: 1018-1026.
- Danilin S, Sourbier C, Thomas L, Rothhut S, Lindner V, Helwig JJ *et al.* von Hippel-Lindau tumor suppressor gene-dependent mRNA stabilization of the survival factor parathyroid hormone-related protein in human renal cell carcinoma by the RNA-binding protein HuR. *Carcinogenesis* 2009; 30: 387-396.
- Dedio J, Konig P, Wohlfart P, Schroeder C, Kummer W, Muller-Esterl W. NOSIP, a novel modulator of endothelial nitric oxide synthase activity. *FASEB J* 2001; 15: 79-89.
- Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 2008; 320: 784-786.
- Dolt KS, Mishra MK, Karar J, Baig MA, Ahmed Z, Pasha MA. cDNA cloning, gene organization and variant specific expression of HIF-1 alpha in high altitude yak (*Bos grunniens*). *Gene* 2007; 386: 73-80.
- Donadi EA, Castelli EC, Arnaiz-Villena A, Roger M, Rey D, Moreau P. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. *Cell Mol Life Sci* 2011; 68: 369-395.
- Droma Y, Ota M, Hanaoka M, Katsuyama Y, Basnyat B, Neupane P *et al.* Two hypoxia sensor genes and their association with symptoms of acute mountain sickness in Sherpas. *Aviat Space Environ Med* 2008; 79: 1056-1060.
- Duplain H, Sartori C, Lepori M, Egli M, Allemann Y, Nicod P *et al.* Exhaled nitric oxide in high-altitude pulmonary edema: role in the regulation of pulmonary vascular tone and evidence for a role against inflammation. *Am J Respir Crit Care Med* 2000; 162: 221-224.
- Ebert BL, Bunn HF. Regulation of the erythropoietin gene. *Blood* 1999; 94: 1864-1877.
- Ellegren H. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat Genet* 2000; 24: 400-402.

- Ellegren H. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 2004; 5: 435-445.
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I. Identification of a variant associated with adult-type hypolactasia. *Nat Genet* 2002; 30: 233-237.
- Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 2010; 10: 564-567.
- Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995; 12: 921-927.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, Salzano FM *et al.* Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 2008; 82: 583-592.
- Fagundes NJ, Salzano FM, Batzer MA, Deininger PL, Bonatto SL. Worldwide genetic variation at the 3'-UTR region of the LDLR gene: possible influence of natural selection. *Ann Hum Genet* 2005; 69: 389-400.
- Fiori G, Facchini F, Pettener D, Rimondi A, Battistini N, Bedogni G. Relationships between blood pressure, anthropometric characteristics and blood lipids in high- and low-altitude populations from Central Asia. *Ann Hum Biol* 2000; 27: 19-28.
- Fletcher W, Yang Z. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol Biol Evol* 2010; 27: 2257-2267.
- Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997; 147: 915-925.
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics* 1993; 133: 693-709.
- Fuselli S, Gilman RH, Chanock SJ, Bonatto SL, De Stefano G, Evans CA *et al.* Analysis of nucleotide diversity of NAT2 coding region reveals homogeneity across Native American populations and high intra-population diversity. *Pharmacogenomics J* 2007; 7: 144-152.
- Fuselli S, Tarazona-Santos E, Dupanloup I, Soto A, Luiselli D, Pettener D. Mitochondrial DNA diversity in South America and the genetic history of Andean highlanders. *Mol Biol Evol* 2003; 20: 1682-1691.
- Galban S, Martindale JL, Mazan-Mamczarz K, Lopez de Silanes I, Fan J, Wang W *et al.* Influence of the RNA-binding protein HuR in pVHL-regulated p53 expression in renal carcinoma cells. *Mol Cell Biol* 2003; 23: 7083-7095.
- Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 1994; 11: 725-736.
- Goldstein DB. Common genetic variation and human traits. *N Engl J Med* 2009; 360: 1696-1698.
- Gonzales GF, Gasco M, Tapia V, Gonzales-Castaneda C. High serum testosterone levels are associated with excessive erythrocytosis of chronic mountain sickness in men. *Am J Physiol Endocrinol Metab* 2009; 296: E1319-1325.
- Grocott M, Montgomery H, Vercueil A. High-altitude physiology and pathophysiology: implications and relevance for intensive care medicine. *Crit Care* 2007; 11: 203.
- Groves BM, Droma T, Sutton JR, McCullough RG, McCullough RE, Zhuang J *et al.* Minimal hypoxic pulmonary hypertension in normal Tibetans at 3,658 m. *J Appl Physiol* 1993; 74: 312-318.
- Hagg M, Wennstrom S. Activation of hypoxia-induced transcription in normoxia. *Exp Cell Res* 2005; 306: 180-191.
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G *et al.* Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet* 2008; 4: e32.
- Harding RM, Healy E, Ray AJ, Ellis NS, Flanagan N, Todd C *et al.* Evidence for variable selective pressures at MC1R. *Am J Hum Genet* 2000; 66: 1351-1361.
- Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 2006; 22: 2971-2972.

- Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* 1994; 91: 9700-9704.
- Hochachka PW, Rupert JL. Fine tuning the HIF-1 'global' O₂ sensor for hypobaric hypoxia in Andean high-altitude natives. *Bioessays* 2003; 25: 515-519.
- Hoit BD, Dalton ND, Erzurum SC, Laskowski D, Strohl KP, Beall CM. Nitric oxide and cardiopulmonary hemodynamics in Tibetan highlanders. *J Appl Physiol* 2005; 99: 1796-1801.
- Holmquist-Mengelbier L, Fredlund E, Lofstedt T, Noguera R, Navarro S, Nilsson H *et al.* Recruitment of HIF-1alpha and HIF-2alpha to common target genes is differentially regulated in neuroblastoma: HIF-2alpha promotes an aggressive phenotype. *Cancer Cell* 2006; 10: 413-423.
- Hu CJ, Iyer S, Sataur A, Covello KL, Chodosh LA, Simon MC. Differential regulation of the transcriptional activities of hypoxia-inducible factor 1 alpha (HIF-1alpha) and HIF-2alpha in stem cells. *Mol Cell Biol* 2006; 26: 3514-3526.
- Hu CJ, Wang LY, Chodosh LA, Keith B, Simon MC. Differential roles of hypoxia-inducible factor 1alpha (HIF-1alpha) and HIF-2alpha in hypoxic gene regulation. *Mol Cell Biol* 2003; 23: 9361-9374.
- Huicho L, Xing G, Qualls C, Rivera-Ch M, Gamboa JL, Verma A *et al.* Abnormal energy regulation in early life: childhood gene expression may predict subsequent chronic mountain sickness. *BMC Pediatr* 2008; 8: 47.
- Iliopoulos O, Ohh M, Kaelin WG, Jr. pVHL19 is a biologically active product of the von Hippel-Lindau gene arising from internal translation initiation. *Proc Natl Acad Sci U S A* 1998; 95: 11661-11666.
- Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000; 408: 708-713.
- Jafarifar F, Yao P, Eswarappa SM, Fox PL. Repression of VEGFA by CA-rich element-binding microRNAs is modulated by hnRNP L. *EMBO J* 2011;
- Jelkmann W. Erythropoietin after a century of research: younger than ever. *Eur J Haematol* 2007; 78: 183-205.
- Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009; 37: D412-416.
- Jobling M, Hurles M, Tyler-Smith C. Human Evolutionary Genetics. *Garland Science* 2004a;
- Jobling MA, Hurles ME, Tyler-Smith C. Human Evolutionary Genetics: origins, peoples and disease. *London/New York: Garland Science Publishing* 2004b;
- Kaelin WG, Jr. The von Hippel-Lindau tumour suppressor protein: O₂ sensing and cancer. *Nat Rev Cancer* 2008; 8: 865-873.
- Kaelin WG, Jr., Ratcliffe PJ. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* 2008; 30: 393-402.
- Kent WJ, Hsu F, Karolchik D, Kuhn RM, Clawson H, Trumbower H *et al.* Exploring relationships and mining data with the UCSC Gene Sorter. *Genome Res* 2005; 15: 737-741.
- Kim U, Wooding S, Ricci D, Jorde LB, Drayna D. Worldwide haplotype diversity and coding sequence variation at human bitter taste receptor loci. *Hum Mutat* 2005; 26: 199-204.
- Kim WY, Kaelin WG. Role of VHL gene mutation in human cancer. *J Clin Oncol* 2004; 22: 4991-5004.
- Kimura M. Evolutionary rate at the molecular level. *Nature* 1968; 217: 624-626.
- Kimura M. The Neutral Theory of Molecular Evolution. *New York: Cambridge University Press* 1983: 367 pp.
- Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* 1971; 68: 820-823.

- Koritzinsky M, Magagnin MG, van den Beucken T, Seigneuric R, Savelkoul K, Dostie J *et al.* Gene expression during acute and prolonged hypoxia is regulated by distinct mechanisms of translational control. *EMBO J* 2006; 25: 1114-1125.
- Kreitman M, Di Rienzo A. Balancing claims for balancing selection. *Trends Genet* 2004; 20: 300-304.
- LaJevic MD, Koduvayur SP, Caffrey V, Cohen RL, Chambers DA. Thy-1 mRNA destabilization by norepinephrine a 3' UTR cAMP responsive decay element and involves RNA binding proteins. *Brain Behav Immun* 2010; 24: 1078-1088.
- Latif F, Tory K, Gnarr J, Yao M, Duh FM, Orcutt ML *et al.* Identification of the von Hippel-Lindau disease tumor suppressor gene. *Science* 1993; 260: 1317-1320.
- Lendahl U, Lee KL, Yang H, Poellinger L. Generating specificity and diversity in the transcriptional response to hypoxia. *Nat Rev Genet* 2009; 10: 821-832.
- Leon-Velarde F, Mejia O. Gene expression in chronic high altitude diseases. *High Alt Med Biol* 2008; 9: 130-139.
- Leslie DMj, Schaller GB. *Bos grunniens* and *Bos mutus* (Artiodactyla: Bovidae). *MAMMALIAN SPECIES* 2009; 836:
- Lewinsky RH, Jensen TG, Moller J, Stensballe A, Olsen J, Troelsen JT. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. *Hum Mol Genet* 2005; 14: 3945-3953.
- Lewontin RC. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 1964; 49: 49-67.
- Lewontin RC, Kojima K. The evolutionary dynamics of complex polymorphisms. *Evolution* 1960; 14: 450-472.
- Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; 319: 1100-1104.
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; 25: 1451-1452.
- Liu N, Gong K, Na X, Wu G, Na YQ. [Analysis of two single nucleotide polymorphisms in von Hippel-Lindau gene and detection of loss of heterozygosity in Chinese sporadic renal cell carcinoma]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2005; 22: 82-84.
- Louicharoen C, Patin E, Paul R, Nuchprayoon I, Witoonpanich B, Peerapittayamongkol C *et al.* Positively selected G6PD-Mahidol mutation reduces Plasmodium vivax density in Southeast Asians. *Science* 2009; 326: 1546-1549.
- Loytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* 2005; 102: 10557-10562.
- Loytynoja A, Goldman N. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 2008; 320: 1632-1635.
- Macdonald SJ, Long AD. Prospects for identifying functional variation across the genome. *Proc Natl Acad Sci U S A* 2005; 102 Suppl 1: 6614-6621.
- Majmundar AJ, Wong WJ, Simon MC. Hypoxia-inducible factors and the response to hypoxic stress. *Mol Cell* 2010; 40: 294-309.
- Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008; 9: 387-402.
- Mattevi A. To be or not to be an oxidase: challenging the oxygen reactivity of flavoenzymes. *Trends Biochem Sci* 2006; 31: 276-283.
- Matthews BH. Effects of High Altitude on Man. *Br Med J* 1945; 2: 75-78.
- McRonald FE, Morris MR, Gentle D, Winchester L, Baban D, Ragoussis J *et al.* CpG methylation profiling in VHL related and VHL unrelated renal cell carcinoma. *Mol Cancer* 2009; 8: 31.
- Mengoni Goñalons GL. Camelids in ancient Andean societies: A review of the zooarchaeological evidence. *Quaternary International* 2008; 185:

- 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; 152: 1103-1110.
- Michel T, Vanhoutte PM. Cellular signaling and NO production. *Pflugers Arch* 2010; 459: 807-816.
- Monge C. Life in the Andes and Chronic Mountain Sickness. *Science* 1942; 95: 79-84.
- Moore LG, Zamudio S, Zhuang J, Droma T, Shohet RV. Analysis of the myoglobin gene in Tibetans living at high altitude. *High Alt Med Biol* 2002; 3: 39-47.
- Morpurgo G, Arese P, Bosia A, Pescarmona GP, Luzzana M, Modiano G *et al.* Sherpas living permanently at high altitude: a new pattern of adaptation. *Proc Natl Acad Sci U S A* 1976; 73: 747-751.
- Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York 1987.
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986; 3: 418-426.
- Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. Oxford University Press 2000.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; 42: 30-35.
- Nielsen R. Statistical tests of selective neutrality in the age of genomics. *Heredity* 2001; 86: 641-647.
- Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet* 2005; 39: 197-218.
- Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat Rev Genet* 2007; 8: 857-868.
- Nordstrom-O'Brien M, van der Looij RB, van Rooijen E, van den Ouweland AM, Majoor-Krakauer DF, Lolkema MP *et al.* Genetic analysis of von Hippel-Lindau disease. *Hum Mutat* 2010; 31: 521-537.
- Pakendorf B, Stoneking M. Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet* 2005; 6: 165-183.
- Paliege A, Rosenberger C, Bondke A, Sciesielski L, Shina A, Heyman SN *et al.* Hypoxia-inducible factor-2 α -expressing interstitial fibroblasts are the only renal cells that express erythropoietin under hypoxia-inducible factor stabilization. *Kidney Int* 2010; 77: 312-318.
- Palmer BF. Physiology and pathophysiology with ascent to altitude. *Am J Med Sci* 2010; 340: 69-77.
- Penalzoza D, Arias-Stella J. The heart and pulmonary circulation at high altitudes: healthy highlanders and chronic mountain sickness. *Circulation* 2007; 115: 1132-1146.
- Percy MJ, McMullin MF, Lappin TR. Sequence analysis of the 3' hypoxia-responsive element of the human erythropoietin gene in patients with erythrocytosis. *Biochem Mol Med* 1997; 62: 132-134.
- Prugnolle F, Manica A, Balloux F. Geography predicts neutral genetic diversity of human populations. *Curr Biol* 2005; 15: R159-160.
- Przeworski M, Hudson RR, Di Rienzo A. Adjusting the focus on human variation. *Trends Genet* 2000; 16: 296-302.
- Rajput C, Najib S, Norboo T, Afrin F, Qadar Pasha MA. Endothelin-1 gene variants and levels associate with adaptation to hypobaric hypoxia in high-altitude natives. *Biochem Biophys Res Commun* 2006; 341: 1218-1224.
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A* 2005; 102: 15942-15947.
- Rankin EB, Biju MP, Liu Q, Unger TL, Rha J, Johnson RS *et al.* Hypoxia-inducible factor-2 (HIF-2) regulates hepatic erythropoietin in vivo. *J Clin Invest* 2007; 117: 1068-1077.
- Raval RR, Lau KW, Tran MG, Sowter HM, Mandriota SJ, Li JL *et al.* Contrasting properties of hypoxia-inducible factor 1 (HIF-1) and HIF-2 in von Hippel-Lindau-associated renal cell carcinoma. *Mol Cell Biol* 2005; 25: 5675-5686.

- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA *et al.* Genetic structure of human populations. *Science* 2002; 298: 2381-2385.
- Rothhammer F, Dillehay TD. The late Pleistocene colonization of South America: an interdisciplinary perspective. *Ann Hum Genet* 2009; 73: 540-549.
- Rupert JL, Devine DV, Monsalve MV, Hochachka PW. Beta-fibrinogen allele frequencies in Peruvian Quechua, a high-altitude native population. *Am J Phys Anthropol* 1999; 109: 181-186.
- Rupert JL, Kidd KK, Norman LE, Monsalve MV, Hochachka PW, Devine DV. Genetic polymorphisms in the Renin-Angiotensin system in high-altitude and low-altitude Native American populations. *Ann Hum Genet* 2003; 67: 17-25.
- Rupert JL, Monsalve MV, Devine DV, Hochachka PW. Beta2-adrenergic receptor allele frequencies in the Quechua, a high altitude native population. *Ann Hum Genet* 2000; 64: 135-143.
- Rytkonen KT, Williams TA, Renshaw GM, Primmer CR, Nikinmaa M. Molecular Evolution of the Metazoan PHD-HIF Oxygen-Sensing System. *Mol Biol Evol* 2011:
- Rzhetsky A, Iossifov I, Koike T, Krauthammer M, Kra P, Morris M *et al.* GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J Biomed Inform* 2004; 37: 43-53.
- Saunders MA, Hammer MF, Nachman MW. Nucleotide variability at G6pd and the signature of malarial selection in humans. *Genetics* 2002; 162: 1849-1861.
- Sawka MN, Young AJ, Rock PB, Lyons TP, Boushel R, Freund BJ *et al.* Altitude acclimatization and blood volume: effects of exogenous erythrocyte volume expansion. *J Appl Physiol* 1996; 81: 636-642.
- Scheinfeldt LB, Soi S, Tishkoff SA. Colloquium paper: working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci U S A* 2010; 107 Suppl 2: 8931-8938.
- Schurr TG. THE PEOPLING OF THE NEW WORLD: Perspectives from Molecular Anthropology. *Annual Review of Anthropology* 2004; 33: 551-583.
- Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriwether DA, Lawrence DN *et al.* Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. *Am J Hum Genet* 1990; 46: 613-623.
- Schurr TG, Sherry ST. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* 2004; 16: 420-439.
- Semenza GL. Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends Mol Med* 2001; 7: 345-350.
- Semenza GL. Hydroxylation of HIF-1: oxygen sensing at the molecular level. *Physiology (Bethesda)* 2004; 19: 176-182.
- Semenza GL. Life with oxygen. *Science* 2007a; 318: 62-64.
- Semenza GL. Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1. *Biochem J* 2007b; 405: 1-9.
- Semenza GL. Involvement of oxygen-sensing pathways in physiologic and pathologic erythropoiesis. *Blood* 2009; 114: 2015-2019.
- Semenza GL, Neufeldt MK, Chi SM, Antonarakis SE. Hypoxia-inducible nuclear factors bind to an enhancer element located 3' to the human erythropoietin gene. *Proc Natl Acad Sci U S A* 1991; 88: 5680-5684.
- Semenza GL, Wang GL. A nuclear factor induced by hypoxia via de novo protein synthesis binds to the human erythropoietin gene enhancer at a site required for transcriptional activation. *Mol Cell Biol* 1992; 12: 5447-5454.
- Shams I, Avivi A, Nevo E. Hypoxic stress tolerance of the blind subterranean mole rat: expression of erythropoietin and hypoxia-inducible factor 1 alpha. *Proc Natl Acad Sci U S A* 2004; 101: 9698-9703.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008; 26: 1135-1145.

- Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Science* 2010; 329: 72-75.
- Smith TG, Brooks JT, Balanos GM, Lappin TR, Layton DM, Leedham DL *et al.* Mutation of von Hippel-Lindau tumour suppressor and human cardiopulmonary physiology. *PLoS Med* 2006; 3: e290.
- Smith TG, Robbins PA, Ratcliffe PJ. The human side of hypoxia-inducible factor. *Br J Haematol* 2008; 141: 325-334.
- Sokal RR, Rohlf FJ. Biometry. 2^o edition WH Freeman and Co, San Francisco, CA 1981:
- Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 2009; 136: 731-745.
- Soranzo N, Bufe B, Sabeti PC, Wilson JF, Weale ME, Marguerie R *et al.* Positive selection on a high-sensitivity allele of the human bitter-taste receptor TAS2R16. *Curr Biol* 2005; 15: 1257-1265.
- Stanish C. The Origin of State Society in South America. *AnnuRevAnthropol* 2001; 30: 41-64.
- StatSoft. StatSoft srl 2005. STATISTICA (sistema software di analisi dei dati) versione 7.1. www.statsoft.it.
- Stickle NH, Cheng LS, Watson IR, Alon N, Malkin D, Irwin MS *et al.* Expression of p53 in renal carcinoma cells is independent of pVHL. *Mutat Res* 2005; 578: 23-32.
- Stobdan T, Karar J, Pasha MA. High altitude adaptation: genetic perspectives. *High Alt Med Biol* 2008; 9: 140-147.
- Strohl KP. Lessons in hypoxic adaptation from high-altitude populations. *Sleep Breath* 2008; 12: 115-121.
- Sturm RA. Human pigmentation genes and their response to solar UV radiation. *Mutat Res* 1998; 422: 69-76.
- Sturm RA, Box NF, Ramsay M. Human pigmentation genetics: the difference is only skin deep. *Bioessays* 1998; 20: 712-721.
- Sufan RI, Moriyama EH, Mariampillai A, Roche O, Evans AJ, Alajez NM *et al.* Oxygen-independent degradation of HIF-alpha via bioengineered VHL tumour suppressor complex. *EMBO Mol Med* 2009; 1: 66-78.
- Tajima F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983; 105: 437-460.
- Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; 123: 585-595.
- Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993; 10: 512-526.
- Tanimoto K, Makino Y, Pereira T, Poellinger L. Mechanism of regulation of the hypoxia-inducible factor-1 alpha by the von Hippel-Lindau tumor suppressor protein. *EMBO J* 2000; 19: 4298-4309.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM *et al.* Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 2001; 68: 1485-1496.
- Tarazona-Santos E, Lavine M, Pastor S, Fiori G, Pettener D. Hematological and pulmonary responses to high altitude in Quechuas: a multivariate approach. *Am J Phys Anthropol* 2000; 111: 165-176.
- Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS *et al.* Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* 2007; 39: 31-40.
- Tishkoff SA, Varkanyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001; 293: 455-462.

- Tong Z, Yang Z, Patel S, Chen H, Gibbs D, Yang X *et al.* Promoter polymorphism of the erythropoietin gene in severe diabetic eye and kidney complications. *Proc Natl Acad Sci U S A* 2008; 105: 6998-7003.
- Tranchevent LC, Barriot R, Yu S, Van Vooren S, Van Loo P, Coessens B *et al.* ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res* 2008; 36: W377-384.
- Valdivia E. Mountain sickness. *Am J Nurs* 1961; 61: 77.
- van den Beucken T, Koritzinsky M, Wouters BG. Translational control of gene expression during hypoxia. *Cancer Biol Ther* 2006; 5: 749-755.
- Wang DP, Li HG, Li YJ, Guo SC, Yang J, Qi DL *et al.* Hypoxia-inducible factor 1alpha cDNA cloning and its mRNA and protein tissue specific expression in domestic yak (*Bos grunniens*) from Qinghai-Tibetan plateau. *Biochem Biophys Res Commun* 2006; 348: 310-319.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G *et al.* Genetic variation and population structure in native Americans. *PLoS Genet* 2007; 3: e185.
- Watterson GA. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* 1975; 7: 256-276.
- Wen D, Boissel JP, Tracy TE, Gruninger RH, Mulcahy LS, Czelusniak J *et al.* Erythropoietin structure-function relationships: high degree of sequence homology among mammals. *Blood* 1993; 82: 1507-1516.
- Wheeler JC, Chikhi L, Bruford MW. Genetic Analysis of the Origins of Domestic South American Camelids. *Zender, MA (Ed), Documenting domestication: New genetic and Archaeological Paradigm University of California Press, Berkeley* 2006; chapter 23: 331-343.
- Whitlock MC, Schluter D. Analisi statistica dei dati biologici. Edizione italiana a cura di Giorgio Bertorelle. *ZANICHELLI* 2010:
- Williams D. Adaptation and acclimatisation in humans and animals at high altitude. *Thorax* 1994; 49 Suppl: S9-13.
- Wilson JN, Rockett K, Keating B, Jallow M, Pinder M, Sisay-Joof F *et al.* A hallmark of balancing selection is present at the promoter region of interleukin 10. *Genes Immun* 2006; 7: 680-683.
- Wooding S, Kim UK, Bamshad MJ, Larsen J, Jorde LB, Drayna D. Natural selection and molecular evolution in PTC, a bitter-taste receptor gene. *Am J Hum Genet* 2004; 74: 637-646.
- Wright S. Genetical structure of populations. *Nature* 1950; 166: 247-249.
- Wrischnik LA, Higuchi RG, Stoneking M, Erlich HA, Arnheim N, Wilson AC. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. *Nucleic Acids Res* 1987; 15: 529-542.
- Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997; 13: 555-556.
- Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 1998; 15: 568-573.
- Yang Z. *Computational Molecular Evolution*. Oxford University Press, OSEE; 2006.
- Yang Z. Adaptive molecular evolution. Handbook of statistical genetics, Chapter 12. *Balding, Bishop and Cannings, John Wiley & Sons, Ltd* 2007a; Volume 1:
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007b; 24: 1586-1591.
- Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* 2000; 15: 496-503.
- Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 2005; 22: 1107-1118.
- Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 2010; 329: 75-78.
- Young AC, Craven RA, Cohen D, Taylor C, Booth C, Harnden P *et al.* Analysis of VHL Gene Alterations and their Relationship to Clinical Parameters in Sporadic Conventional Renal Cell Carcinoma. *Clin Cancer Res* 2009; 15: 7582-7592.



OG-250
Features & Benefits

Primary packaging

- Genomic DNA equivalent to DNA from blood for downstream applications
- Compact, robust design is ideal for storage and mailing
- Non-invasive, reliable self-collection that increases compliance and decreases costs
- Long term specimen stability at room temperature
- Higher quality and greater than 15x the DNA yield collected with buccal swabs
- Purification reagent available for easy purification
- Optional accessories available for collection from children and non-spitters
- FDA registered, CE certified

Product Specifications

OG-250 (in outer shell)
 Dimensions: 8.0 x 12.0 x 2.5cm
 Weight: 30.3 g

Sample ready for transport/storage



Instructions

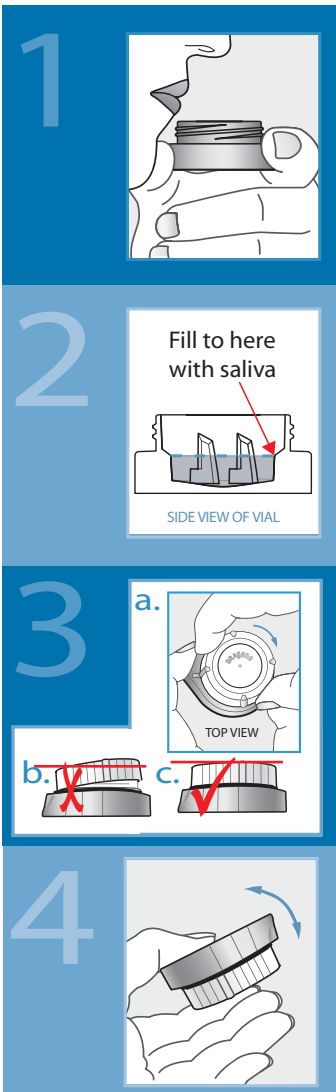
Cap



Base



Appendice 1



ENGLISH Oragene™ DNA Self-Collection Kit User Instructions

Do not eat, drink, smoke or chew gum for 30 minutes before giving your saliva sample.

- 1 Start giving your saliva sample.
- 2 Spit until the amount of liquid saliva (not bubbles) reaches at least the level shown in picture #2.
- 3 Put the container on a flat surface. Screw the cap onto the container. Make sure that the cap is closed tightly, as shown in picture #3c. If you have trouble closing the cap all the way, turn the cap slightly counter-clockwise then try to close it again.
- 4 Mix gently for at least 10 seconds.
Recycle outer packaging.

User Tips:

- Do NOT remove plastic film from the cap.
- To make more saliva, close your mouth and wiggle your tongue or rub your cheeks. Some people find it easier to spit more if they place ¼ tsp of plain white sugar on your tongue.
- Finish spitting within 30 minutes.

Intended Use: This product is designed for the safe collection of saliva samples.

Contents: Each collection vial contains 2ml of Oragene liquid. Before use, the solution in the cap should be clear and colorless.

Warnings: Wash with water if the Oragene liquid comes in contact with eyes or skin. Do not ingest.

Storage: Store at room temperature 15-30 °C (59-86 °F).

FRANÇAIS Oragene™ : kit de prélèvement d'ADN Notice d'utilisation

Il est demandé de ne pas manger, boire, fumer ou mâcher du chewing-gum pendant les 30 minutes qui précèdent le prélèvement de l'échantillon.

- 1 Recueillez l'échantillon de salive.
- 2 Salivez jusqu'à ce que le liquide (sans bulles) atteigne le niveau indiqué sur la photo n° 2.
- 3 Disposez le flacon sur une surface plane. Vissez le bouchon du flacon. Assurez-vous que le bouchon est correctement vissé, comme indiqué sur la photo n° 3c. Si vous ne parvenez pas à visser correctement le bouchon, dévissez-le légèrement avant d'essayer de le visser à nouveau.
- 4 Mélangez doucement pendant au moins 10 secondes.
Recyclez l'emballage.

Conseils d'utilisation :

- N'enlevez PAS la pellicule plastique du bouchon.
- Pour produire plus de salive, fermez votre bouche et remuez la langue ou passez-la sur les parois internes de vos joues. Certaines personnes produisent plus de salive en plaçant ¼ de cuillère à thé de sucre blanc sur la langue.
- Recueillez l'échantillon dans un délai de 30 minutes.

Utilisation du produit : ce produit a été conçu pour prélever des échantillons d'ADN par la salive en toute sécurité.

Contenu : chaque flacon contient 2 ml de solution Oragene liquide. Celle-ci doit être claire et incolore avant l'utilisation.

Avertissement : si la solution Oragene entre en contact avec les yeux ou la peau, rincez avec de l'eau. Ne pas ingérer.

Entreposage : gardez le flacon à température ambiante, entre 15 et 30 °C (59-86 °F).

ESPAÑOL Kit de recolección de ADN Oragene™ Instrucciones para el usuario

No coma, beba, fume ni mastique goma de mascar en los 30 minutos previos a entregar su muestra de saliva.

- 1 Comience a dar su muestra de saliva.
- 2 Llene el recipiente con saliva líquida (sin contar la espuma) hasta alcanzar o superar el límite que aparece en la figura número 2.
- 3 Coloque el contenedor en una superficie plana. Asegure la tapa al contenedor. Asegúrese de que la tapa esté bien cerrada, tal como se muestra en la figura número 3c. Si tiene problemas para cerrar la tapa por completo, de vuelta a la tapa ligeramente, en sentido contrario a las agujas del reloj, y luego inténtelo de nuevo.
- 4 Mezcle ligeramente durante al menos 10 segundos.
Recicle la envoltura externa.

Sugerencias para el usuario:

- NO quite la película de plástico de la tapa.
- Para producir más saliva, cierre la boca y mueva la lengua o frótese las mejillas. Algunas personas encuentran más fácil escupir más si colocan ¼ de cucharadita de azúcar blanca común sobre la lengua.
- Termine de recolectar la muestra en un plazo de 30 minutos.

Uso específico: Este producto está diseñado para la recolección segura de muestras de saliva.

Contenido: Cada frasco contiene 2 ml de líquido Oragene. Antes de utilizarla, la solución en la tapa debe ser transparente e incolora.

Advertencia: Si el líquido Oragene entra en contacto con los ojos o la piel, lave con agua. No lo ingiera.

Almacenamiento: Guárdese a temperatura ambiente entre 15-30 °C (59-86 °F).

DEUTSCH Oragene™ DNA-Selbstentnahme-Set Gebrauchsanweisung

Stellen Sie 30 Minuten vor Entnahme der Speichelprobe Essen, Trinken, Rauchen und Kaugummikauen ein.

- 1 Beginnen Sie mit der Entnahme Ihrer Speichelprobe.
- 2 Geben Sie so viel Speichel ab, dass die Menge an flüssigem Speichel (nicht Luftblasen) den in Abb. 2 angegebenen Füllstand erreicht oder überschreitet.
- 3 Stellen Sie den Behälter auf eine ebene Oberfläche. Schrauben Sie die Kappe auf den Behälter. Stellen Sie sicher, dass der Behälter fest verschlossen ist (wie in Abb. 3c dargestellt). Sollten Sie den Behälter nicht komplett schließen können, drehen Sie die Kappe etwas gegen den Uhrzeigersinn und versuchen Sie erneut, ihn zu schließen.
- 4 Schwenken Sie mindestens 10 Sekunden lang vorsichtig den Behälter.
Entsorgen Sie die Verpackung.

Gebrauchshinweise:

- Entfernen Sie NICHT die Kunststoffolie von der Kappe.
- Um mehr Speichel zu erzeugen, bewegen Sie bei geschlossenem Mund die Zunge oder reiben Sie Ihre Wangen. Die Speichelabgabe wird erleichtert, wenn Sie 1/4 Teelöffel reinen weißen Zucker auf Ihre Zunge geben.
- Die Speichelabgabe sollte innerhalb von 30 Min. abgeschlossen sein.

Verwendungszweck: Dieses Produkt dient zur sicheren Entnahme von Speichelproben.

Inhalt: Jeder Entnahmebehälter enthält 2 ml Oragene-Flüssigkeit. Vor Verwendung muss die Lösung im Behälter klar und farblos sein.

Warnung: Wenn die Oragene-Flüssigkeit mit Augen oder Haut in Berührung kommt, mit Wasser spülen. Flüssigkeit nicht einnehmen.

Lagerung: Bei Raumtemperatur (15-30 °C) lagern.

NEDERLANDS Oragene™ doe-het-zelfkit voor het verzamelen van DNA Gebruiksaanwijzing

U mag in het halfru vooraf uw speekselmonster afstaan niet eten, drinken of roken en geen kaugumg kauwen.

- 1 Sta een speekselmonster af.
- 2 Blijf spugen tot de hoeveelheid vloeibaar speeksel (schuim telt niet) het in afbeelding 2 aangegeven niveau bereikt.
- 3 Plaats de container op een vlakke ondergrond. Schroef de dop op de container. Schroef de dop zoals aangegeven in afbeelding 3c goed vast. Draai als het u niet lukt de dop helemaal dicht te schroeven de dop een stukje terug (linksom). Draai de dop vervolgens opnieuw vast.
- 4 Schud de container gedurende minimaal 10 seconden.
Recycle de buitenverpakking.

Tips voor de gebruiker:

- Verwijder de plastic tape NIET van de dop.
- Maak meer speeksel aan door met uw mond dicht uw tong te bewegen of langs uw wang te wrijven. Het is gemakkelijker om meer te spugen als u ¼ theelep kristalsuiker op uw tong legt.
- U dient de container binnen 30 minuten te vullen.

Bedoeld gebruik: dit product is bedoeld voor het op veilige wijze verzamelen van speekselmonsters.

Inhoud: elk verzamelbuisje bevat 2 ml Oragene-vloeistof. De vloeistof in de dop dient voorafgaand aan het gebruik helder en kleurloos te zijn.

Waarschuwing: als de Oragene-vloeistof met uw ogen of huid in contact komt, dient u deze te spoelen met water. Slik de vloeistof niet door.

Bewaren: bewaar dit product bij kamertemperatuur (15 - 30 °C).

SVENSKA Oragene™ - personligt uppsamlings-kit för DNA Användarinstruktioner

Du bör inte äta, dricka, röka eller tugga tuggummi 30 minuter innan du lämnar ditt salivprov.

- 1 Börja lämna salivprovet.
- 2 Spotta tills mängden saliv (inte bubblor) når upp till eller överstiger nivån som visas på bild 2.
- 3 Ställ behållaren på en plan yta. Skruva på korken på behållaren. Se till att korken sluter helt tätt enligt bild 3c. Om du har svårt att skruva åt korken helt kan du vrida den aningen moturs och sedan försöka skruva åt den igen.
- 4 Blanda försiktigt i minst 10 sekunder.
Återvinn den yttre förpackningen.

Tips:

- Ta INTE bort plastfilmen från locket.
- För att kunna utsöndra mer saliv stänger du munnen och rör sedan tungan fram och tillbaka eller gnuggar kinderna. Vissa personer anser att det blir lättare att spotta om man placerar ¼ tsk vanligt vitt socker på tungan.
- Sluta spotta inom 30 minuter.

Avsedd användning: Den här produkten är utformad för att på ett säkert sätt samla in salivprov.

Innehåll: Varje provampull innehåller 2 ml Oragene-vätska. Innan du börjar ska lösningen vara klar och färglös.

Varningar: Tvätta med vatten om Oragene-vätskan kommer i kontakt med ögon eller hud. Svälj inte.

Förvaring: Förvara i rumstemperatur vid 15-30 °C.

简体中文 Oragene™ DNA 自采集工具包 用户说明

在提供唾液样本之前的 30 分钟内不要饮食、抽烟或嚼口香糖。

- 1 开始提供唾液样本。
- 2 请一直吐口水，直到唾液（而不是泡沫）量达到或超过图 2 显示的水平。
- 3 将容器放平。旋紧容器盖。确保容器盖紧闭，如图 3c 所示。如果无法旋紧容器盖，请沿逆时针方向稍微转动盖子，然后再次尝试旋紧。
- 4 请至少轻轻摇晃 10 秒钟。循环利用外包装。

用户提示:

- 不要拿掉盖子上的塑料膜。
- 要得到更多唾液，请合上嘴，摆动舌头，或摩擦脸颊。有些人会在舌头上放 ¼ tsp 的纯白糖时，更容易吐出更多唾液。
- 在 30 分钟内吐好唾液。

用途: 本产品设计用于唾液样本的安全采集。

容量: 每个采集瓶含 2ml 的 Oragene 液体。使用前，盖子中的溶液应清澈无色。

警告: 如果眼睛或皮肤接触到 Oragene 液体，请用水冲洗。请勿咽下。

储存: 储存室温应为 15-30 °C (59-86 °F)。

PD-PR-024 Issue 1.0 OG 250 User Instructions

EC REP
Emergo Europe
Molenstraat 15, 2513 BH The Hague
The Netherlands
Tel: (+31) (0) 70 345-8570
Fax: (+31) (0) 70 346-7299

Label Legend
Consult package insert
In Vitro Medical Device
CE Marking
Collect saliva by

Légende de l'étiquette
Consultez l'information à l'intérieur
Produit médical in vitro
Marque CE
Prélevez la salive avant le

Leyenda de la etiqueta
Lea las instrucciones del paquete
Dispositivo médico de diagnóstico in vitro
Marca CE
Para recolectar saliva

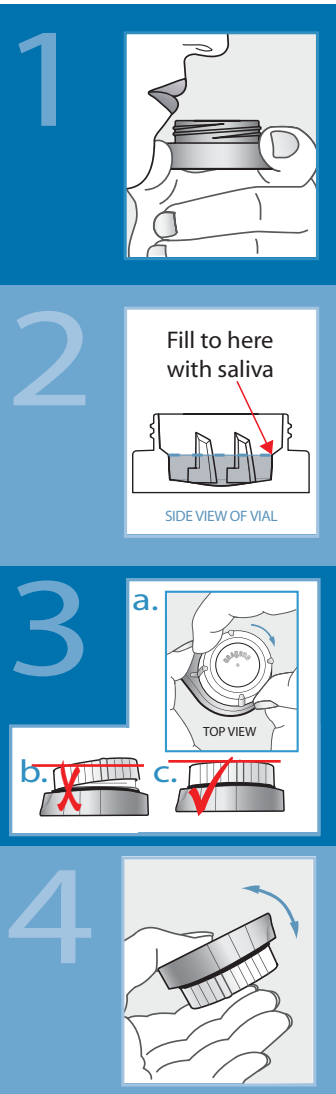
Legende des Etiketts
Packungsbeilage beachten
In-vitro-Diagnostikum
CE-Kennzeichnung
Speichelentnahme bis

Uitleg van etiket
Zie verpakingsbijsluiters
In vitro medisch hulpmiddel
CE-merk
Speeksel verzamelen voor

Förklaring av etikett
Se bruksanvisning i förpackningen
Medicinsk enhet för in vitro-användning
CE-märkning
Samla in saliv före

标签图例
查看药品说明书
体外医疗设备
CE 标记
唾液采集方法

Appendice 1



SUOMI
Oragene™ DNA:n keräysastia
Käyttöohje

Älä syö, juo, polta tupakkaa tai pure purukumia 30 minuuttia ennen sylkinäytteen antamista.

- 1 Aloita sylkinäytteen antaminen.
- 2 Sylje niin paljon, että sylkinesteen (vaahtoa ei lasketa) pinta on vähintään kuvassa 2 näkyvällä tasolla.
- 3 Aseta purkki tasaiselle pinnalle. Kierrä kansi kiinni purkkiin. Varmista, että kansi on suljettu tiiviisti kuvassa 3c näkyvällä tavalla. Jos kansi ei sulkeudu kunnolla, kierrä sitä hieman vastapäivään ja yritä sulkea se uudelleen.
- 4 Sekoita varovasti vähintään 10 sekunnin ajan.

Kierrätä tuotteen pakkaus.

Vinkkejä:

- ÄLÄ irrota kannen muovikalvoa.
- Voit lisätä syljen eritystä pitämällä suuta kiinni ja liikuttamalla kieltä tai hieromalla poskia. Joillakin ihmisillä syljen erityis lisääntyy, kun kielen päälle asetetaan ¼ tl tavallista hienoa sokeria.
- Käytä näytteen antamiseen korkeintaan 30 minuuttia.

Käyttötarkoitus: Tämä tuote on suunniteltu sylkinäytteiden turvalliseen keräämiseen.

Sisältö: Kussakin keräysastiasa on 2 ml Oragene-nestettä. Ennen käyttöä keräysastiasa olevan liuoksen on oltava kirkasta ja väritöntä.

Varoitukset: Jos Oragene-nestettä joutuu iholle tai silmiin, huuhtelee alue huolellisesti vedellä. Ei saa niellä.

Säilytys: Säilytys huoneenlämmössä 15–30 °C.

DANSK
Oragene™ DNA Self-Collection Kit
Brugsanvisning

Du må ikke spise, drikke, ryge eller tygge tyggegummi 30 minutter, før du afgiver din spytp prøve.

- 1 Afgiv din spytp prøve.
- 2 Spyt indtil mængden af flydende spyt (uden bobler) når eller overstiger det niveau, der vises på billedet nr. 2.
- 3 Anbring beholderen på et plant underlag. Skru hættten på beholderen. Kontroller, at hættten er lukket korrekt som vist på billedet nummer 3c. Hvis du har problemer med at lukke hættten helt, skal du dreje hættten en anelse mod uret og derefter forsøge at lukke den igen.
- 4 Ryst forsigtigt i mindst 10 sekunder.

Send yderemballagen til genbrug.

Bruertips:

- Fjern IKKE plastfilmen fra hættten.
- For at danne mere spyt kan du lukke munden og bevæge tungen eller gnide dig på kinderne. Det kan være lettere at producere spyt, hvis du lægger ¼ tsk. sukker på tungen.
- Du skal være færdig med at spytte inden for 30 minutter.

Tilsigtet anvendelse: Dette produkt er beregnet til sikker indsamling af spytp prøve

Indhold: Hver opsamlingsbeholder indeholder 2 ml Oragene-væske. Før brug skal opløsningen i hættten være klar og farveløs.

Advarsel: Skyl med vand, hvis du får Oragene-væske på huden eller i øjnene. Må ikke indtages.

Opbevaring: Opbevares ved stuetemperatur 15-30 °C.

ITALIANO
Kit di autoraccolta della saliva Oragene™
per il test del DNA
Istruzioni

Non mangiare, bere, fumare o masticare chewing gum per 30 minuti prima della raccolta del campione di saliva.

- 1 Emettere la saliva.
- 2 Riempire il contenitore di saliva liquida (non schiuma) fino a raggiungere o superare il livello indicato nella figura 2.
- 3 Posizionare il contenitore su una superficie piana. Avvitare il tappo del contenitore e assicurarsi che sia chiuso saldamente come mostrato nella figura 3c. Se non si riesce ad avvitare completamente il tappo, girarlo in senso antiorario, quindi riprovare.
- 4 Agitare con cura per almeno 10 secondi.

Riciclare la confezione esterna.

Consigli per l'uso:

- NON rimuovere la pellicola di plastica dal tappo.
- Per produrre più saliva, chiudere la bocca e muovere la lingua oppure massaggiare le guance. In alternativa, mettere sulla lingua ¼ cucchiaino di zucchero bianco.
- Completare la raccolta del campione entro 30 minuti.

Indicazioni: questo prodotto è stato ideato per la raccolta sicura di campioni di saliva.

Contenuto: ciascun contenitore di raccolta contiene 2 ml di liquido Oragene. Prima dell'uso, la soluzione nel tappo deve presentarsi limpida e incolore.

Attenzione: nel caso in cui il prodotto entri a contatto con gli occhi, lavare con acqua. Non ingerire.

Conservazione: conservare a una temperatura ambiente di 15-30 °C.

NORSK
Oragene™ DNA-prøvetakingssett
Brukerveiledning

Ikke spis, drikk, røyk eller tygg tyggegummi de siste 30 minuttene før du avgir spyttprøven.

- 1 Begynn å avgi spyttprøven.
- 2 Fortsett å samle opp spytt til mengden flytende spytt (ikke bobler) når opp til eller overstiger nivået som er vist på bilde 2.
- 3 Sett beholderen på en jevn overflate. Skru lokket på beholderen. Pass på at lokket er godt skrudd igjen, som vist på bilde 3c. Hvis det er vanskelig å skru lokket helt igjen, kan du skru lokket litt mot klokken og deretter prøve å skru det igjen på nytt.
- 4 Bland prøven forsiktig i minst 10 sekunder.

Lever inn ytteremballasjen til gjenvinning.

Brukertips:

- IKKE fjern plastfilmen fra lokket.
- Hvis du trenger å produsere mer spytt, kan du lukke munnen og bevege på tungen eller gni deg i kindene. Noen synes det blir lettere å produsere nok spytt hvis du legger ¼ ts hvitt sukker på tungen.
- Spyttprøven må tas i løpet av 30 minutter.

Bruksområde: Dette produktet skal brukes til å ta sikre spyttprøver.

Innhold: Hvert prøveglass inneholder 2 ml Oragene-væske. Før bruk skal løsningen være klar og fargeløs.

Advarsel: Skyll med vann hvis Oragene-væsken kommer i kontakt med øynene eller huden. Må ikke svelges.

Oppbevaring: Oppbevares ved romtemperatur, 15–30 °C (59–86 °F).

PORTUGUÊS
Kit de Auto-Recolha de ADN Oragene™
Instruções para o Utilizador

Não coma, beba, fume ou mastigue pastilha elástica na meia-hora anterior à recolha de saliva.

- 1 Inicie o processo de recolha.
- 2 Cuspa até que a quantidade de saliva líquida (e não espuma) atinja ou ultrapasse o nível indicado na imagem 2.
- 3 Coloque o recipiente numa superfície plana. Enrosque a tampa no recipiente. Certifique-se que a tampa se encontra bem apertada, como demonstrado na imagem 3c. Se não conseguir fechar o recipiente totalmente, gire a tampa para a esquerda e tente novamente.
- 4 Misture suavemente durante cerca de 10 segundos.

Recicle o invólucro exterior.

Sugestões:

- NÃO retire a película de plástico da tampa.
- Para produzir mais saliva, com a boca fechada, agite a língua ou esfregue as bochechas. Quando colocam um ¼ de uma colher de sobremesa de açúcar na boca, algumas pessoas conseguem cuspir uma maior quantidade de saliva com mais facilidade.
- Acabe de cuspir num período de 30 minutos.

Utilização: Este produto destina-se à recolha, em segurança, de amostras de saliva.

Conteúdo: Cada frasco de recolha contém 2 ml de líquido Oragene. Antes de utilizar, a solução na tampa deve ser transparente e incolor.

Aviso: Se o líquido Oragene entrar em contacto com a pele ou com os olhos, lave com água. Não ingerir.

Armazenamento: Guarde à temperatura ambiente de 15-30°C.

ΕΛΛΗΝΙΚΑ
Kit ατομικής συλλογής δειγμάτων DNA Oragene™
Οδηγίες χρήσης

Μην τρώτε, μην πίνετε, μην καπνίζετε και μη μασάτε τσίχλα για 30 λεπτά προτού δώσετε δείγμα από το σάλιο σας.

- 1 Ξεκινήστε να δίνετε δείγμα από το σάλιο σας.
- 2 Φτύνετε έως ότου η ποσότητα του υγρού σάλιου (χωρίς τον αφρό) φτάσει ή ξεπεράσει τη στάθμη που υποδεικνύεται στην εικόνα 2.
- 3 Τοποθετήστε τον περιέκτη σε επίπεδη επιφάνεια. Βιδώστε το πώμα στον περιέκτη. Βεβαιωθείτε ότι το πώμα έχει κλείσει καλά, όπως υποδεικνύεται στην εικόνα 3c. Εάν δυσκολεύεστε να κλείσετε τελείως το πώμα, στρέψτε το ελαφρώς αριστερόστροφα και στη συνέχεια προσπαθήστε να το ξανακλείσετε.
- 4 Ανακινήστε με απαλές κινήσεις για τουλάχιστον 10 δευτερόλεπτα.

Ανακυκλώστε την εξωτερική συσκευασία.

Χρήσιμες συμβουλές:

- ΜΗΝ αφαιρείτε την πλαστική μεμβράνη από το πώμα.
- Για την παραγωγή επιπλέον ποσότητας σάλιου, κλείστε το στόμα σας και κινήστε τη γλώσσα σας ή τρίψτε απαλά τα μάγουλά σας. Ορισμένα άτομα το βρίσκουν ευκολότερο να φτύσουν μεγαλύτερη ποσότητα σάλιου, εάν τοποθετήσουν ¼ του κουταλιού του γλυκού απλή λευκή ζάχαρη επάνω στη γλώσσα τους.
- Ολοκληρώστε τη διαδικασία τοποθέτησης σάλιου εντός 30 λεπτών.

Προοριζόμενη χρήση: Το προϊόν αυτό είναι σχεδιασμένο για την ασφαλή συλλογή δειγμάτων σάλιου.

Περιεχόμενο: Κάθε φιαλίδιο συλλογής δειγμάτων περιέχει 2ml υγρού Oragene. Προ της χρήσης, το διάλυμα που εμπεριέχεται στο πώμα θα πρέπει να είναι διαυγές και άχρωμο.

Προειδοποιήσεις: Αν το υγρό Oragene έρθει σε επαφή με το δέρμα ή τα μάτια, ξεπλύνετε με νερό. Μην το καταπίνετε.

Φύλαξη: Φυλάσσεται σε θερμοκρασία δωματίου μεταξύ 15 και 30°C.

日本語
Oragene™ DNA 自己採取キット
取扱説明書

唾液のサンプルを採取する 30 分前から、飲食、喫煙、およびガムを噛むのは控えてください。

- 1 唾液サンプルの採取を開始します。
- 2 唾液の量が（気泡なしで）図 2 のレベル以上になるまで吐き出します。
- 3 容器を平面に置きます。容器にキャップをかぶせて閉めます。図 3c のようにキャップがしっかりと閉まっていることを確認します。キャップが最後まで閉まらない場合は、キャップを少し左に回してゆるめから、もう一度閉め直してみてください。
- 4 10 秒以上静かに容器を振ります。

外装を再利用します。

取扱上の注意:

- キャップのプラスチックフィルムを取り除かないでください。
- 唾液の量を増やすには、口を閉じて舌を動かすか、頬をさすります。舌の上にスプーン小さじ 4 分の 1 ほどの白砂糖を載せると、唾液が出やすくなる場合があります。
- 唾液の採取は 30 分以内で終わらせてください。

使用目的: この製品は、唾液サンプルを安全に採取するためのものです。

内容: 採取バイアルには、Oragene 液が 2ml 入っています。使用前のキャップ内の液体は、透明無色です。

注意: Oragene 液が眼や皮膚に付着した場合は、水で洗い流してください。飲み込まないでください。

保管: 室温（15～30°C）で保管してください。

Laboratory Protocol for Manual Purification of DNA from 0.5 mL of Oragene®•DNA/saliva

DNA yield and stability in Oragene•DNA

Oragene•DNA is designed to capture the large amount of DNA present in saliva. The median yield of DNA from 2 mL of saliva when captured in 2 mL of Oragene•DNA is 110 µg.

When saliva is mixed with the Oragene•DNA solution, the DNA is immediately stabilized. Oragene•DNA/saliva samples are stable at room temperature for years without processing. If it is your laboratory practice to store frozen samples, Oragene•DNA/saliva samples can be stored indefinitely at -15 to -20°C, and can undergo multiple freeze-thaw cycles without deterioration of the DNA.

Manual purification according to the following protocol will produce a high yield of DNA. Alternatively, Oragene•DNA is also compatible with many high throughput DNA purification systems.

The following step-by-step protocol describes how to purify DNA from a 500 µL aliquot of an Oragene•DNA/saliva sample. Volumes of less than 500 µL may be purified by adjusting the volumes of reagents accordingly.

Equipment and reagents to be supplied by user

- Microcentrifuge capable of running at 13,000 rpm (15,000 × g)
- Air or water incubator at 50°C (Note: water incubator is not recommended for OG-500).
- Ethanol (95 to 100%) at room temperature
- DNA buffer: TE (10 mM Tris-HCl, 1mM EDTA, pH 8.0) or similar solution
- (Optional) Glycogen (20 mg/mL) (e.g., Invitrogen Cat. No. 10814-010)
- Ethanol (70%) at room temperature

Procedure

Purification Steps	Notes
1. Mix the Oragene•DNA/saliva sample in the Oragene•DNA vial by inversion and gentle shaking for a few seconds.	<ul style="list-style-type: none"> • This is to ensure that viscous saliva samples are properly mixed with the Oragene•DNA solution.
2. Incubate the sample at 50°C in a water incubator for a minimum of 1 hour or in an air incubator for a minimum of 2 hours.	<ul style="list-style-type: none"> • DNA in Oragene•DNA is stable at room temperature even without the incubation step. • This heat-treatment step is essential to ensure that DNA is adequately released and that nucleases are permanently inactivated. • This incubation step may be performed at any time after saliva is collected and before it is purified. • Incubation of the entire sample is recommended to ensure that this step has been completed. However, if it is more convenient, a 500 µL aliquot can be heat-treated. • The sample can be incubated either in the original container or after transfer to another tube. • The sample may be incubated at 50°C overnight if it is more convenient. • A longer time is required in an air incubator because temperature equilibration is slower than in a water incubator.
3. Transfer 500 µL of the mixed Oragene•DNA/saliva sample to a 1.5 mL microcentrifuge tube.	<ul style="list-style-type: none"> • The remainder of the Oragene•DNA/saliva sample can be stored at room temperature or frozen (-15°C to -20°C).

Purification Steps	Notes
4. For 500 μ L of Oragene•DNA/saliva, add 20 μ L (1/25th volume) of Oragene•DNA Purifier (OG-L2P, supplied) to the microcentrifuge tube and mix by vortexing for a few seconds.	<ul style="list-style-type: none"> The sample will become turbid as impurities and inhibitors are precipitated.
5. Incubate on ice for 10 minutes.	<ul style="list-style-type: none"> Room temperature incubation can be substituted but will be slightly less effective in removing impurities.
6. Centrifuge at room temperature for 5 minutes at 13,000 rpm (15,000 $\times g$).	<ul style="list-style-type: none"> A longer period of centrifugation (up to 15 min) may be beneficial in reducing the turbidity (high A_{320}) of the final DNA solution.
7. Carefully transfer the clear supernatant with a pipet tip into a fresh microcentrifuge tube. Discard the pellet containing impurities. Optional: Addition of Glycogen	<ul style="list-style-type: none"> The pellet contains turbid impurities. If accidentally disturbed, the tube should be re-centrifuged. <p><i>Optional: Addition of Glycogen</i></p> <ul style="list-style-type: none"> Some users may prefer to add 5 μL (100 μg) of Glycogen to the supernatant to make the pellet more easily visible.
8. To 500 μ L of supernatant, add 500 μ L (i.e., an equal volume) of room-temperature 95-100% ethanol. Mix gently by inversion 10 times.	<ul style="list-style-type: none"> During mixing with ethanol, the DNA will be precipitated. This may appear as a clot of DNA fibers or as a fine precipitate, depending upon the amount of DNA in the sample. Even if no clot is seen, DNA will be recovered by carefully following the next steps.
9. Allow the sample to stand at room temperature for 10 minutes to allow the DNA to fully precipitate.	<ul style="list-style-type: none"> Incubation at -20°C is not recommended because impurities may co-precipitate with the DNA.
10. Place the tube in the microcentrifuge in a known orientation. Centrifuge at room temperature for 2 minutes at 13,000 rpm (15,000 $\times g$).	<ul style="list-style-type: none"> For example, place each tube in the microcentrifuge with the hinge portion of the cap pointing away from the centre of the rotor. After centrifugation, the position of the pellet can be located (even if too tiny to be easily visible.): it will be at the tip of the tube below the hinge.

Purification Steps	Notes
11. Carefully remove the supernatant with a pipet tip and discard it. Take care to avoid disturbing the DNA pellet.	<ul style="list-style-type: none"> • This pellet contains DNA. Loss of the pellet will result in loss of the DNA. • Rotating the tube such that the pellet is on the upper wall will allow you to safely move a pipet tip along the lower wall and remove all of the supernatant. • The supernatant may contain impurities and should be removed as completely as possible. • Excessive drying of the pellet can make the DNA more difficult to dissolve.
12. Ethanol wash step Carefully add 250 μ L of 70% ethanol. Let stand at room temperature for 1 minute. Completely remove the ethanol without disturbing the pellet.	<ul style="list-style-type: none"> • Take care not to disturb the DNA pellet. • The DNA pellet may be small. Addition of a carrier such as glycogen at step #7 will increase the visibility of the pellet. • Should the pellet detach, centrifuge the sample for 5 minutes at 13,000 rpm (15,000 x g). • The 70% ethanol wash helps to remove residual inhibitors.
13. Add 100 μ L of DNA buffer (see Reagents) to dissolve the DNA pellet. Vortex for at least 5 seconds.	<ul style="list-style-type: none"> • The expected concentration of the fully hydrated DNA is 20 to 200 ng/μL. • Note that large amounts of high molecular weight DNA can be slow to hydrate (dissolve) completely. • Incomplete hydration of the DNA is a cause of inaccuracy in estimating DNA concentration and of failure of downstream applications such as PCR.
14. (Optional) Additional steps to ensure complete hydration of the DNA.	<p>a) Additional vigorous pipetting and vortexing, and/or</p> <p>b) Incubation at 50°C for 1 hour with occasional vortexing, and/or</p> <p>c) Incubation at room temperature for 1-2 days</p> <ul style="list-style-type: none"> • For applications such as Southern blotting that require very high molecular weight DNA, (c) is recommended.
15. Storage of the fully rehydrated DNA.	<ul style="list-style-type: none"> • In TE at 4°C for up to 1-2 months. • Recommended in TE in aliquots at -20°C for long-term storage. • Note that freezing of purified DNA in TE will cause DNA to precipitate. When thawing a sample of frozen purified DNA, pay careful attention to rehydration, as discussed in step 13.

Quantification of DNA:

Assays that use fluorescent dyes are more specific than absorbance at 260nm for quantifying the amount of double-stranded DNA (dsDNA) in a saliva DNA sample. We recommend using fluorescent dyes such as Picogreen or Sybrgreen to quantify dsDNA since there is less interference by contaminating RNA. An inexpensive protocol using Sybrgreen can be found on our website at [http://www.dnagenotek.com/pdf_files/PD-PR-075Issue1.0_DNAQuantificationusingSYBRGreen1.pdf.] Alternatively, commercially available kits such as Invitrogen's Quant-iT™ Pico Green® dsDNA Assay Kit Cat. No. Q-33130 can be used. For either protocol, we recommend that the purified DNA be diluted 1:50 with TE buffer and that 5 µL be used in the quantification assay.

Suggestions for quantifying by absorbance:

If you choose to quantify DNA by absorbance, we recommend that you first treat the purified sample with RNase to digest contaminating RNA and then remove the RNA fragments by ethanol precipitation of the DNA. A detailed protocol can be found on our website at [http://www.dnagenotek.com/pdf_files/RNA_removal_by_double-RNase_Digestion_Issue_1.2.pdf.] Please note that DNA from saliva typically contains appreciably more RNA than found in blood samples. Ensure that alcohol-precipitated DNA is fully dissolved before reading the absorbance.

Conversion Factor: an absorbance of 1.0 at 260 nm corresponds to a concentration of 50 ng/µL (50 µg/mL) for pure dsDNA.

- A spectrophotometer cuvette capable of reading a volume of 100 µL or less should be used to avoid using too large a volume of sample.
- Absorbance values at 260 nm should be between 0.1 and 1.5. Lower values may not be reliable. If the undiluted sample is used, care must be taken to ensure that the cuvette is very clean or that disposable cells are used to avoid cross-contamination of samples. Absorbance values >1.5 at 260 nm are not reliable; the sample should be

Method:

1. Dilute a 10 µL aliquot of purified RNase treated DNA with 90 µL of TE (1/10 dilution). Mix by gently pipetting up and down. Wait for bubbles to clear.
2. Use TE in the reference (blank) cell.
3. Measure absorbance at 320 nm, 280 nm and 260 nm.
4. Calculate corrected A_{280} and A_{260} values by subtracting the absorbance at 320 nm (A_{320}) from the A_{280} and A_{260} values.
5. DNA concentration in ng/µL = corrected $A_{260} \times 10$ (dilution factor) $\times 50$ (conversion factor).
6. A_{260}/A_{280} Ratio: Divide corrected A_{260} by corrected A_{280} .

Example:

1. Assume the measured $A_{320} = 0.025$, $A_{280} = 0.175$ and $A_{260} = 0.295$
2. The DNA concentration of the undiluted sample will be
 $(A_{260} - A_{320}) \times 10$ [dilution factor] $\times 50$ [conversion factor]
 $= (0.295 - 0.025) \times 10 \times 50$
 $= 0.270 \times 10 \times 50$
 $= 135 \text{ ng/}\mu\text{L}$ or $135 \text{ }\mu\text{g/mL}$
3. The corrected A_{260}/A_{280} ratio will be
 $(A_{260} - A_{320}) \div (A_{280} - A_{320})$
 $= (0.295 - 0.025) \div (0.175 - 0.025)$
 $= 0.270 \div 0.150$
 $= 1.80$

Appendice 2

ID	Sito di campionamento	Età	Sesso	Gruppo etnico	Altezza (cm)	Peso (Kg)	Perimetro del torace (cm)		
							normale	massimo	minimo
CBA001	Cusibamba	35	F	Quechua	146	63	97	99	94
CBA002	Cusibamba	50	M	Quechua	160	57	89	97	87
CBA003	Cusibamba	36	F	Quechua	154	62	88	92	87
CBA004	Cusibamba	28	F	Quechua	152	55	85	90	84
CBA005	Cusibamba	34	M	Quechua	156	47	89	92	87
CBA006	Cusibamba	52	F	Quechua	151	59	90	93	89
CBA007	Cusibamba	50	F	Quechua	156	52	90	93	85
CBA008	Cusibamba	32	M	Quechua	152	55	95	100	93
CBA009	Cusibamba	35	F	Quechua	143	46	88	91	87
CBA010	Cusibamba	30	F	Quechua	143	46	84	86	82
CBA011	Cusibamba	55	M	Quechua	162	59	90	94	88
CBA012	Cusibamba	29	F	Quechua	160	54	87	93	84
CBA013	Cusibamba	29	F	Quechua	149	50	87	90	83
CBA014	Cusibamba	23	F	Quechua	156	53	81	85	77
CBA015	Cusibamba	47	F	Quechua	150	44	79	83	77
CBA016	Cusibamba	27	M	Quechua	167	75	96	100	93
CBA017	Cusibamba	27	M	Quechua	163	60	88	93	85
CBA018	Cusibamba	73	M	Quechua	153	50	92	94	88
CBA019	Cusibamba	64	F	Quechua	141	36	70	73	70
CBA020	Cusibamba	72	F	Quechua	147	45	80	82	78
CBA021	Cusibamba	61	F	Quechua	154	48	83	87	81
CBA022	Cusibamba	53	F	Quechua	154	50	87	93	86
CBA023	Cusibamba	39	F	Quechua	149	62	93	95	89
CBA024	Cusibamba	34	F	Quechua	155	55	88	93	85
CUS001	UNSAAC	19	F	Quechua	155	45	81	82	78

CUS002	UNSAAC	66	M	Quechua	162	70	112	115	111
CUS003	UNSAAC	19	M	Quechua	165	55	85	87	83
CUS004	UNSAAC	18	F	Quechua	156	53	78	81	77
CUS005	UNSAAC	21	F	Quechua	144	45	81	85	80
CUS006	UNSAAC	20	F	Quechua	154	46	75	79	73
CUS007	UNSAAC	21	M	Quechua	174	61	88	91	80
CUS008	UNSAAC	21	F	Quechua	162	47	77	83	71
CUS009	UNSAAC	37	M	Quechua	171	74	99	105	96
CUS010	UNSAAC	33	M	Quechua	165	68	96	100	94
CUS011	UNSAAC	21	M	Quechua	168	63	90	92	85
CUS012	UNSAAC	28	M	Quechua	160	55	90	95	86
CUS013	UNSAAC	28	M	Mestizo	170	75	95	100	92
CUS014	UNSAAC	25	F	Quechua	152	45	72	74	68
CUS015	UNSAAC	22	F	Quechua	160	55	83	85	78
CUS016	UNSAAC	24	M	Quechua	172	73	94	101	92
CUS017	UNSAAC	26	M	Quechua	176	69	89	92	85
CUS018	UNSAAC	23	M	Quechua	169	61	87	88	82
CUS019	UNSAAC	27	M	Quechua	172	69	91	95	89
CUS020	UNSAAC	24	M	Quechua	153	46	82	85	78
CUS021	UNSAAC	20	F	Quechua	154	48	78	79	74
CUS022	UNSAAC	25	M	Quechua	174	55	88	92	86
CUS023	UNSAAC	21	M	Quechua	169	51	81	86	76
CUS024	UNSAAC	22	M	Quechua	171	64	95	98	92
CUS025	UNSAAC	21	M	Quechua	167	58	94	98	87
CUS026	UNSAAC	22	M	Quechua	163	61	91	94	87
CUS027	UNSAAC	22	M	Quechua	165	62	89	95	87
CUS028	UNSAAC	19	F	Quechua	148	45	76	82	74
CUS029	UNSAAC	18	F	Quechua	155	56	83	89	81
CUS030	UNSAAC	18	M	Quechua	173	56	85	88	81

CUS031	UNSAAC	19	F	Quechua	157	53	80	82	77
CUS032	UNSAAC	23	F	Quechua	161	51	80	83	79
CUS033	UNSAAC	22	F	Quechua	158	53	76	80	74
CUS034	UNSAAC	30	F	Quechua	147	44	79	83	78
CUS035	UNSAAC	23	F	Quechua	155	48	82	85	76
CUS036	UNSAAC	20	M	Quechua	174	62	94	99	89
CUS037	UNSAAC	26	F	Quechua	143	43	78	80	76
CUS038	UNSAAC	23	M	Quechua	171	51	86	88	80
CUS039	UNSAAC	33	F	Quechua	155	52	83	84	79
CUS040	UNSAAC	20	F	Quechua	162	49	80	84	78
CUS041	UNSAAC	22	M	Quechua/Aymara	166	72	95	100	91
CUS042	UNSAAC	21	M	Quechua	184	68	86	92	82
CUS043	UNSAAC	21	F	Quechua	169	51	74	77	69
CUS044	UNSAAC	26	M	Quechua	162	62	90	94	87
CUS045	UNSAAC	21	F	Quechua	161	55	82	91	79
CUS046	UNSAAC	30	M	Quechua	170	66	97	101	91
CUS047	UNSAAC	27	M	Mestizo	170	65	93	97	90
CUS048	UNSAAC	35	M	Espanol	168	60	90	93	86
CUS049	UNSAAC	23	F	Quechua	164	54	77	80	74
CUS050	UNSAAC	22	M	Quechua	170	63	93	95	86
CUS051	UNSAAC	24	F	Quechua	153	49	80	82	79
CUS052	UNSAAC	22	F	Quechua	154	52	87	90	81
CUS053	UNSAAC	28	F	Quechua	157	61	85	88	82
CUS054	UNSAAC	19	F	Mestizo	162	53	77	80	74
CUS055	UNSAAC	29	M	Quechua	175	68	91	95	87
CUS056	UNSAAC	18	F	Quechua	156	61	86	90	85
CUS057	UNSAAC	20	M	Quechua	169	56	85	87	80
CUS058	UNSAAC	21	F	Quechua	162	55	81	82	77
CUS059	UNSAAC	22	M	Quechua	169	86	107	109	105

CUS060	UNSAAC	27	M	Quechua	166	55	87	91	83
CUS061	UNSAAC	29	F	Quechua	159	65	97	98	96
PUN001	Puno - Banco de sangre	42	M	Aymarà	166	65	95	98	91
PUN002	Puno - Banco de sangre	24	M	Quechua	165	74	100	103	97
PUN003	Puno - Banco de sangre	44	F	Quechua	148	56	84	88	83
PUN004	Puno - Banco de sangre	42	F	Quechua	151	56	85	88	84
PUN005	Puno - Banco de sangre	51	M	Quechua	168	80	99	103	97
PUN006	Puno - Banco de sangre	26	F	Quechua	168	70	78	81	76
PUN007	Puno - Banco de sangre	26	M	Quechua	185	65	93	96	91
PUN008	Puno - Banco de sangre	49	M	Aymarà	165	65	105	106	98
PUN009	Puno - Banco de sangre	32	F	Quechua	160	70	88	92	85
PUN010	Puno - Banco de sangre	54	F	Quechua	149	54	89	91	87
PUN011	Puno - Banco de sangre	23	M	Quechua	166	66	90	95	86
PUN012	Puno - Banco de sangre	26	F	Aymarà	155	49	78	81	75
PUN013	Puno - Banco de sangre	24	F	Aymarà	152	56	84	87	79
PUN014	Puno - Banco de sangre	58	F	Aymarà	148	77	105	107	103
PUN015	Puno - Banco de sangre	44	F	Quechua	150	62	92	94	91
PUN016	Puno - Banco de sangre	62	F	Quechua	163	65	98	100	95
PUN017	Puno - Banco de sangre	50	F	Mestizo	167	80	99	101	97
PUN018	Puno - Banco de sangre	61	F	Quechua	150	40	83	87	80
PUN019	Puno - Banco de sangre	56	M	Quechua	175	115	135	139	133
PUN020	Puno - Banco de sangre	54	M	Aymarà	174	104	118	120	117
PUN021	Puno - Banco de sangre	55	M	Quechua	164	84	110	112	108
PUN022	Puno - Banco de sangre	61	M	Quechua	165	68	103	105	100
PUN023	Puno - Banco de sangre	51	M	Quechua - Aymarà	160	75	102	103	101
PUN024	Puno - Banco de sangre	43	M	Quechua	166	78	100	102	98
PUN025	Puno - Banco de sangre	22	F	Quechua	154	55	93	95	91
PUN026	Puno - Banco de sangre	49	F	Mestizo	159	74	108	110	106
PUN027	Puno - Banco de sangre	23	F	Quechua - Aymarà	155	51	80	81	79

PUN028	Puno - Banco de sangre	29	F	Quechua	155	46	84	87	81
PUN029	Puno - Banco de sangre	59	M	Quechua	158	56	94	95	96
PUN030	Puno - Banco de sangre	18	M	Quechua - Aymará	172	69	91	85	88
PUN031	Puno - Banco de sangre	18	F	Mestizo	146	48	88	89	85
PUN032	Puno - Banco de sangre	37	M	Quechua/mestizo	168	78	101	106	98
PUN033	Puno - Banco de sangre	48	M	Aymará	165	75	90	94	88
PUN034	Puno - Banco de sangre	47	M	Quechua	161	65	93	96	91
PUN035	Puno - Banco de sangre	31	F	Quechua	167	84	95	99	94
PUN036	Puno - Banco de sangre	30	M	Mestizo	178	100	106	110	105
PUN037	Puno - Banco de sangre	60	F	Mestizo	156	90	113	115	112
PUN038	Puno - Banco de sangre	56	F	Aymará	158	78	109	110	108
PUN039	Puno - Banco de sangre	55	M	Quechua	174	75	96	101	95
PUN040	Puno - Banco de sangre	53	F	Quechua	158	67	100	99	97
PUN041	Puno - Banco de sangre	23	M	Quechua	175	65	90	94	88
PUN042	Puno - Banco de sangre	46	F	Aymará	154	80	106	107	104
PUN043	Puno - Banco de sangre	41	F	Quechua	148	62	96	98	93
PUN044	Puno - Banco de sangre	53	F	Quechua	152	55	94	95	93
PUN045	Puno - Banco de sangre	36	F	Aymará	161	92	117	118	116
PUN046	Puno - Banco de sangre	32	F	Aymará	170	65	94	96	92
PUN047	Puno - Banco de sangre	30	M	Quechua	168	74	99	113	96
PUN048	Puno - Banco de sangre	28	M	Quechua	169	68	95	98	93
PUN049	Puno - Banco de sangre	30	M	Quechua	173	84	104	106	102
PUN050	Puno - Banco de sangre	32	M	Quechua	169	82	100	97	104
PUN051	Puno - Banco de sangre	23	F	Aymará	164	59	90	88	92
PUN052	Puno - Banco de sangre	29	F	Quechua	152	73	109	110	107
PUN053	Puno - Banco de sangre	24	M	Quechua	173	67	86	92	84
PUN054	Puno - Banco de sangre	20	F	Aymará	152	44	76	81	74
PUN055	Puno - Banco de sangre	35	M	Quechua	170	100	106	109	104
PUN056	Puno - Banco de sangre	36	F	Quechua-Aymara	161	80	91	94	88

PUN057	Puno - Banco de sangre	50	M	Aymarà	167	70	100	104	98
PUN058	Puno - Banco de sangre	23	M	Quechua-Aymara	174	76	94	97	92
PUN059	Puno - Banco de sangre	57	M	Mestizo	168	90	112	113	110
PUN060	Puno - Banco de sangre	28	M	Quechua	167	75	102	105	99
PUN061	Puno - Banco de sangre	18	F	Aymarà	156	55	77	81	76
PUN062	Puno - Banco de sangre	26	F	Mestizo	165	57	81	84	78
PUN063	Puno - Banco de sangre	26	F	Quechua	159	54	77	81	75
PUN064	Puno - Banco de sangre	37	F	Quechua	165	64	90	93	88
PUN065	Puno - Banco de sangre	35	F	Quechua	154	67	97	99	95
PUN066	Puno - Banco de sangre	57	F	Quechua-Aymara	160	102	111	115	109
PUN067	Puno - Banco de sangre	21	F	Aymarà-Quechua	168	65	87	93	84
PUN068	Puno - Banco de sangre	18	F	Aymarà-Quechua	154	60	82	84	80
PUN069	Puno - Banco de sangre	19	F	Aymarà-Quechua	159	53	73	83	70
PUN070	Puno - Banco de sangre	19	F	Aymarà	156	45	69	65	72
PUN071	Puno - Banco de sangre	18	F	Aymarà-Quechua	160	60	78	83	76
PUN072	Puno - Banco de sangre	21	F	Aymarà-Quechua	154	46	74	77	69
PUN073	Puno - Banco de sangre	21	F	Aymarà	148	54	83	88	81
PUN074	Puno - Banco de sangre	32	F	Quechua	153	45	79	83	77
PUN075	Puno - Banco de sangre	21	F	Quechua	168	61	86	89	85
PUN076	Puno - Banco de sangre	26	F	Aymarà-Quechua	160	60	86	90	83
PUN077	Puno - Banco de sangre	19	F	Mestizo	157	44	72	79	70
PUN078	Puno - Banco de sangre	18	F	Aymarà-Quechua	162	69	80	83	79
PUN079	Puno - Banco de sangre	20	F	Aymarà-Quechua	161	70	89	93	87
PUN080	Puno - Banco de sangre	18	F	Aymarà	157	50	75	77	74
PUN081	Puno - Banco de sangre	49	F	Aymarà	153	74	93	98	91
PUN082	Puno - Banco de sangre	19	F	Aymarà	149	70	92	93	90
PUN083	Puno - Banco de sangre	34	M	Aymarà	173	68	91	97	89
PUN084	Puno - Banco de sangre	24	F	Aymarà-Quechua	155	55	78	82	75
PUN085	Puno - Banco de sangre	21	F	Quechua	149	53	85	87	84

PUN086	Puno - Banco de sangre	33	F	Quechua	148	65	91	93	89
PUN087	Puno - Banco de sangre	80	M	Quechua	164	74	100	104	99
PUN088	Puno - Banco de sangre	32	F	Quechua	153	71	104	106	100
PUN089	Puno - Banco de sangre	50	F	Quechua	152	90	109	111	107
PUN090	Puno - Banco de sangre	20	F	Quechua	147	44	74	77	72
PUN091	Puno - Banco de sangre	21	F	Quechua	161	54	74	80	72
PUN092	Puno - Banco de sangre	19	F	Aymarà	155	64	89	93	87
PUN093	Puno - Banco de sangre	36	M	Quechua	158	79	109	111	107
PUN094	Puno - Banco de sangre	61	F	Quechua	146	65	108	111	104
PUN095	Puno - Banco de sangre	18	F	Mestizo	155	55	79	80	76
PUN096	Puno - Banco de sangre	19	F	Quechua	149	50	74	78	71
PUN097	Puno - Banco de sangre	19	F	Aymarà	151	58	86	88	85
PUN098	Puno - Banco de sangre	20	F	Quechua	147	54	81	85	77
PUN099	Puno - Banco de sangre	18	F	Aymarà	150	78	103	105	101
PUN100	Puno - Banco de sangre	34	M	Quechua	161	63	94	97	91
PUN101	Puno - Banco de sangre	22	F	Aymarà	158	58	87	90	84
PUN102	Puno - Banco de sangre	52	F	Aymarà-Quechua	152	65	96	98	93
PUN103	Puno - Banco de sangre	18	F	Quechua	149	64	91	93	90
PUN104	Puno - Banco de sangre	20	F	Aymarà	156	49	75	78	73
PUN105	Puno - Banco de sangre	53	F	Quechua	149	75	104	107	103
PUN106	Puno - Banco de sangre	65	F	Aymarà	154	66	100	102	96
PUN107	Puno - Banco de sangre	18	F	Aymarà-Quechua	151	45	76	78	73
PUN108	Puno - Banco de sangre	49	F	Mestizo	157	94	118	120	117
PUN109	Puno - Banco de sangre	20	F	Quechua	150	60	94	96	92
PUN110	Puno - Banco de sangre	28	F	Quechua	162	65	91	94	86
PUN111	Puno - Banco de sangre	30	M	Quechua	167	73	98	104	96
PUN112	Puno - Banco de sangre	34	F	Aymarà	154	70	100	102	98
PUN113	Puno - Banco de sangre	28	F	Quechua	159	64	90	92	88
PUN114	Puno - Banco de sangre	30	F	Quechua	159	65	88	90	86

PUN115	Puno - Banco de sangre	62	F	Aymarà	152	59	90	93	87
PUN116	Puno - Banco de sangre	40	F	Mestizo	156	63	93	95	91
PUN117	Puno - Banco de sangre	51	F	Castellano	156	64	87	91	86
PUN118	Puno - Banco de sangre	48	F	Quechua	151	60	77	80	75
PUN119	Puno - Banco de sangre	27	F	Aymarà	164	63	88	92	84
PUN120	Puno - Banco de sangre	29	M	Quechua	171	63	91	96	87
PUN121	Puno - Banco de sangre	42	F	Quechua	147	69	99	100	97
PUN122	Puno - Banco de sangre	19	F	Aymarà	152	54	78	81	76
PUN123	Puno - Banco de sangre	64	F	Aymarà-Quechua	145	56	91	93	86
PUN124	Puno - Banco de sangre	60	M	Aymarà	160	73	95	97	94
PUN125	Puno - Banco de sangre	20	F	Quechua	170	60	84	88	83
PUN126	Puno - Banco de sangre	49	F	Mestizo	149	57	87	89	85
PUN127	Puno - Banco de sangre	22	F	Quechua	149	47	81	85	79
PUN128	Puno - Banco de sangre	22	F	Quechua	157	56	86	88	83
PUN129	Puno - Banco de sangre	22	F	Quechua	146	35	68	70	67
PUN130	Puno - Banco de sangre	21	F	Aymarà	160	56	78	81	76
PUN131	Puno - Banco de sangre	26	F	Aymarà	158	53	78	84	76
PUN132	Puno - Banco de sangre	22	F	Aymarà	154	64	92	96	91
PUN133	Puno - Banco de sangre	20	F	Quechua	156	48	79	81	76
PUN134	Puno - Banco de sangre	21	F	Mestizo	152	50	81	84	77
PUN135	Puno - Banco de sangre	21	F	Quechua	149	49	76	80	74
PUN136	Puno - Banco de sangre	24	F	Aymarà	149	49	80	84	79
PUN137	Puno - Banco de sangre	22	F	Aymarà	160	66	88	91	84
PUN138	Puno - Banco de sangre	22	F	Quechua	150	62	85	89	85
PUN139	Puno - Banco de sangre	21	M	Quechua	163	57	89	91	83
PUN140	Puno - Banco de sangre	26	F	Aymarà-Quechua	158	73	100	102	99
PUN141	Puno - Banco de sangre	24	M	Quechua	165	58	89	92	88
PUN142	Puno - Banco de sangre	23	M	Aymarà	168	63	83	91	81

Pressione sistolica (mmHg) media	Pressione diastolica (mmHg) media	Battito cardiaco (BPM) media	SpO2 (%) media	Temperatura corporea °C	[Hb] g/dl
111	72	66	91	34,2	17,5
109	65	63	92	35,8	16,5
103	62	83	95	36,8	18,2
105	64	77	91	35,9	17,4
103	68	62	98	36,2	17,3
128	63	61	88	35,8	16,5
122	71	71	91	33,5	16,0
128	74	56	94	36,1	19,0
100	64	80	93	34,0	16,9
112	67	73	93	36,1	16,9
128	78	58	94	33,0	18,6
103	59	63	92	34,0	17,6
121	48	61	94	34,8	15,0
89	62	66	92	35,8	14,6
106	69	69	96	33,2	15,5
108	70	74	95	32,7	21,0
111	63	60	70	33,9	19,9
142	82	64	89	33,1	20,0
97	68	78	92	35,9	18,7
104	60	62	90	37,4	16,0
89	55	68	76	32,9	14,6
114	68	67	85	33,5	16,5
113	75	75	85	34,6	17,2
137	82	73	89	33,6	17,6
99	71	72	94	34,3	15,2

121	76	68	92	36,1	18,2
106	63	57	94	36,1	16,5
101	64	66	91	36,3	16,9
129	96	70	91	37,2	14,2
90	56	67	87	33,8	13,2
114	65	57	91	36,2	17,3
113	72	72	84	36,2	18,1
130	87	77	98	37,2	18,5
117	65	57	-	36,7	15,9
127	70	62	96	36,3	14,9
100	49	60	96	36,2	18,3
144	87	75	97	33,5	18,4
112	78	74	93	36,0	17,8
110	78	76	89	36,3	16,9
132	73	73	96	36,3	17,5
128	88	67	-	34,0	20,4
118	67	92	93	35,8	18,0
126	69	79	-	36,8	18,6
131	93	61	90	36,1	19,0
102	64	77	90	36,7	14,9
116	60	76	95	36,3	14,8
130	69	77	93	36,6	15,8
126	62	67	93	36,6	18,3
112	72	87	97	36,8	20,2
127	72	81	93	37,2	16,7
122	60	69	94	37,3	18,3
112	82	67	92	36,7	18,3
108	77	82	95	36,6	18,0
121	61	83	95	37,1	18,6

118	72	78	94	36,1	16,4
106	65	82	96	34,8	17,3
105	65	55	92	1,5	17,6
90	59	87	-	36,3	17,4
104	64	70	94	36,1	17,5
109	64	67	96	36,2	17,4
99	66	65	90	35,8	17,4
97	56	62	97	36,1	18,0
96	65	81	95	36,9	16,4
105	70	83	97	32,9	18,2
119	63	54	91	36,8	18,0
101	61	73	92	36,1	18,0
102	68	87	73	37,1	18,7
105	68	72	95	36,2	17,9
102	61	75	95	37,3	11,5
109	67	69	94	36,2	18,6
112	70	74	92	36,1	18,5
97	54	63	93	36,2	18,4
113	66	70	-	36,7	17,3
119	71	79	95	37,2	17,3
106	70	72	94	36,9	18,0
135	82	80	94	36,6	16,8
115	75	74	-	36,1	14,6
103	67	80	96	35,9	16,2
137	79	78	96	37,2	16,9
109	63	76	94	36,1	17,5
106	53	76	95	36,1	18,3
91	57	78	97	35,9	18,8
134	82	57	98	36,7	18,4

129	75	76	94	36,8	18,0
102	72	83	97	33,9	17,5
119	79	60	86	36,7	17,0
127	78	75	94	36,6	19,0
92	67	76	93	36,0	14,7
101	60	55	92	36,7	14,9
113	75	69	94	36,4	16,8
134	92	73	93	36,8	15,1
126	77	72	94	36,7	17,4
122	76	55	98	35,8	16,6
113	72	74	92	36,1	16,6
118	75	89	93	36,4	13,6
116	75	73	93	36,4	17,3
99	55	71	96	37,0	11,5
109	80	85	88	36,5	14,3
117	70	65	78	35,7	15,6
116	80	60	83	35,7	15,3
126	76	68	92	33,0	15,4
115	70	63	93	35,7	16,8
102	51	56	94	36,7	13,1
141	92	95	93	37,7	14,0
120	88	86	94	35,9	19,6
142	89	75	93	36,0	16,3
113	71	64	98	36,4	17,6
123	85	60	98	34,3	16,9
153	105	71	96	37,1	16,7
105	64	68	93	35,9	16,2
104	77	87	96	36,0	15,1
108	69	75	92	37,0	15,6

122	75	83	93	34,1	16,9
108	85	108	85	35,6	23,4
125	76	68	88	36,1	17,1
119	94	75	94	37,8	17,3
103	75	54	95	36,0	17,0
111	78	69	96	36,2	17,4
89	60	54	95	36,1	17,2
100	57	66	89	36,7	17,8
115	66	91	95	34,5	22,1
119	97	69	91	34,7	19,7
111	71	77	92	36,7	16,4
125	86	71	93	34,8	17,1
110	67	64	94	34,6	15,4
116	72	64	91	36,2	16,4
100	53	66	92	35,9	16,4
118	80	77	89	32,6	17,5
85	58	84	91	36,7	
111	81	95	92	34,2	14,5
106	74	74	92	36,6	14,2
117	75	61	94	36,2	18,3
128	82	74	94	36,8	18,3
109	63	81	93	36,1	17,3
133	79	95	93	37,1	19,3
100	71	74	93	36,0	16,0
119	68	75	97	36,0	17,6
103	68	73	93	37,1	17,3
95	64	71	91	36,6	16,4
102	62	69	95	34,2	21,0
113	83	98	94	36,0	

128	78	71	93	33,6	14,8
114	69	65	94	36,0	14,6
155	104	81	85	36,2	23,2
128	75	93	92	36,6	16,2
85	54	71	94	37,1	16,9
115	79	67	95	34,6	16,2
103	68	73	94	34,2	16,8
114	72	80	93	35,9	13,4
110	71	63	91	36,0	17,1
126	98	80	95	36,5	19,6
88	55	64	92	34,3	16,6
97	59	64	92	36,5	16,1
114	68	61	93	36,6	
116	82	87	92	36,9	14,6
108	69	65	95	34,1	16,8
104	70	75	89	34,5	16,1
110	77	72	74	36,0	17,6
98	64	84	96	34,6	17,1
117	78	73	93	36,5	16,6
107	81	70	95	36,5	16,8
114	78	89	82	36,9	17,0
125	78	71	91	35,8	16,7
107	75	74	95	35,9	17,1
142	105	73	93	36,4	13,8
114	74	63	95	36,6	17,6
120	83	94	92	36,9	16,6
124	70	54	96	36,0	17,8
113	67	75	90	36,4	15,6
103	65	90	89	35,9	15,4

113	79	65	94	35,9	
142	78	60	92	34,0	18,1
98	68	65	78	33,9	16,7
115	71	71	87	36,0	23,1
103	63	57	93	35,8	15,4
100	66	71	85	34,2	17,8
114	71	69	94	33,0	14,9
122	94	61	95	35,9	17,0
98	60	67	97	35,8	17,4
101	71	95	83	37,0	16,9
110	72	74	90	35,9	16,2
99	57	68	93	36,0	16,3
103	69	66	97	33,4	16,0
126	73	67	92	36,5	17,3
113	71	58	-	34,0	20,6
106	71	63	93	37,0	16,8
103	68	66	93	36,8	16,6
130	76	62	88	36,6	16,9
100	58	68	98	35,8	17,9
125	79	78	93	35,5	17,9
115	74	86	96	33,0	16,8
97	51	66	98	35,9	17,0
147	90	75	85	37,0	16,6
97	67	64	97	35,9	16,8
90	64	72	91	36,0	16,6
106	59	59	98	35,8	17,2
101	66	75	93	37,0	17,0
106	62	55	94	36,0	17,0
100	59	65	93	36,4	16,6

129	76	62	97	36,7	14,5
101	67	83	96	36,4	16,1
94	73	58	95	36,8	16,6
99	61	77	95	35,9	16,9
104	71	62	95	35,9	16,3
98	54	62	94	35,7	17,4
170	90	83	91	36,8	20,4
98	59	81	65	36,6	17,0
101	61	72	93	35,7	16,8
120	75	74	96	35,8	19,8
92	61	71	93	36,8	16,9
121	89	75	93	36,8	16,9
97	72	84	92	36,6	15,3
105	67	71	96	36,5	15,7
94	64	79	93	35,8	16,6
102	68	82	92	36,3	15,2
92	58	73	83	35,6	16,4
112	74	76	96	35,9	16,8
117	75	76	93	36,5	16,4
107	72	89	90	36,9	17,8
90	64	71	92	35,9	17,0
96	71	85	93	35,9	16,2
109	60	69	91	35,8	16,7
99	61	69	87	36,9	16,5
114	65	71	97	36,0	17,8
113	71	67	88	34,5	16,9
109	71	113	93	37,1	17,7
93	64	74	92	36,3	17,0

Ematocrito (%)	Gruppo sanguigno	Fattore Rh	Glucosio (mg/dl)	mtDNA Aplogruppo	VHL				
					rs779804	rs779805	rs1642742	rs1681669	rs1681668
53	0	+		B	CC	GG	GG	GG	CC
50	0	+		B	TT	AG	AG	AG	TC
54	0	+		B	CT	GG	GG	GG	CC
52	0	+		A	CT	GG	GG	GG	CC
52	0	+		D	TT	AG	AG	AG	TC
50	0	+		D	CT	GG	GG	GG	CC
49	0	+		D	CT	GG	GG	GG	CC
58	0	+		B	CC	GG	GG	GG	CC
50	0	+		B	TT	AG	AG	AG	TC
51	0	+		C	CT	GG	GG	GG	CC
57	0	+		A	TT	AA	AA	AA	TT
52	0	+		A	CT	GG	GG	GG	CC
45	0	+		B	TT	AG	AG	AG	TC
44	0	+		C	CT	GA	GA	GA	CT
45	0	+		B	TT	AG	AG	AG	TC
66	0	+		A	CT	GG	GG	GG	CC
60	0	+		D	CC	GG	GG	GG	CC
62	0	+		A	TT	AG	AG	AG	TC
57	0	+		B	CT	GG	GG	GG	CC
48	0	+		B	TT	AA	AA	AA	TT
44	0	+		A	CC	GG	GA	GA	CT
48	0	+		D	CT	GA	GA	GA	CT
52	0	+		D	TT	AG	AG	AG	TC
52	0	+		B	TC	AG	AG	AG	TC
53	0	+		B	TT	AA	AA	AA	TT

50	0	+		A					
51	0	+		B	TT	AA	AA	AA	TT
48	0	+		D					
47	0	+		C	TT	AG	AG	AG	TC
49	0	+		B	CT	GA	GA	GA	CT
59	0	+		B					
45	0	+		B	CT	GA	GA	GA	CT
50	A	+		B					
48	0	+		B	TT	AG	AG	AG	TC
56	0	+		B	CT	GA	GA	GA	CT
56	0	+		B					
55	0	+		B					
52	A	+		A	CT	GA	GA	GA	CT
51	AB	+		C					
56	0	+		B					
60	0	+		B					
58	0	+		B					
51	0	+		B					
59	0	+		B					
46	B	+		D	CT	GG	GG	GG	CC
48	0	+		C					
60	0	+		B					
51	0	+		B	TT	AA	AA	AA	TT
58	0	+		C					
50	0	+		B	CT	GG	GG	GG	CC
56	0	+		B	CT	GA	GA	GA	CT
53	0	+		B	TT	AG	GA	AG	TC
57	A	+		?	TT	AA	GG	GG	CC
60	0	+		B					

44	0	+		B					
50	0	+		B					
43	0	+		B					
58	A	+		C	TT	AA	AA	AA	TT
49	A	+		B	CT	GG	GG	GG	CC
52	0	+		C					
46	0	+		C	TT	AA	AA	AA	TT
53	0	+		B					
40	A	+		B					
49	0	+		B	TT	AG	GA	AG	TC
52	0	+		C					
54	0	+		C					
54	A	+		D	CT	GA	GA	GA	CT
53	0	+		B					
33	0	+		B	CT	GA	GA	GA	CT
51	0	+		B	TT	AG	AG	AG	TC
52	A	+		C					
54	0	+		D					
53	0	+		B					
47	0	+		A					
47	A	+		B	CT	GG	GG	GG	CC
49	0	+		B					
42	0	+		B	CC	GG	GG	GG	CC
48	B	+		B					
50	0	+		B					
52	0	+		B	TT	AA	AA	AA	TT
54	A	+		A	TT	AG	AG	AG	TC
58	0	+		D	CT	GA	GA	GA	CT
57	B	+		C	CT	GA	GA	GA	CT

56	0	+		B					
51	0	+		B	CT	GG	GG	GG	CC
51	0	+	63,3	B	CT	GG	GG	GG	CC
58	0	+	65,6	B	CT	GA	GA	GA	CT
43	0	+	70,3	B	TT	AA	AA	AA	TT
45	0	+	61,8	B	CT	GA	GA	GA	CT
50	0	+	64,6	D	CT	GG	GG	GG	CC
44	B	+	70,8	B	TT	GG	GG	GG	CC
53	A	+	63,2	B					
47	0	+	61,6	B	TT	GG	GG	GG	CC
49	0	+	66,4	B					
39	0	+	70,4	B	TT	AA	AA	AA	TT
53	0	+	66,9	B					
33	0	+	65,7	B	CT	GA	GA	GA	CT
44	0	+	60,0	C	TT	AA	AA	AA	TT
45	A	+	68,2	A					
45	0	+	62,7	B	CT	GA	GA	GA	CT
46	0	+	63,4	C	CT	GA	GA	GA	CT
50	0	+	67,6	B					
38	AB	+	62,0	B	CT	GA	GA	GA	CT
43	0	+	71,8	A					
58	0	+	69,6	B	TT	AG	AG	AG	TC
48	0	+	180,1	C					
50	0	+	66,8	C	TT	AG	AG	AG	TC
52	0	+	68,9	B					
48	0	+	64,6	A	TT	AG	AG	AG	TC
48	0	+	57,0	B					
44	B	+	61,0	D					
45	0	+	58,0	A					

48	0	+	67,0	B						
70	0	+	63,0	?	TT	AA	AA	AA	TT	
50	0	+	78,0	C						
54	0	+	67,0	B						
51	0	+	71	B						
53	0	+	69,0	B	CT	GA	GA	GA	CT	
54	0	+	73,1	D						
54	0	-	60,0	B	TT	AG	AG	AG	TC	
68	B	-	61,0	B						
60	B	+	60,0	B						
50	0	+	300,0	A						
52	0	+	65,0	B	CT	GA	GA	GA	CT	
47	0	+	59,0	B	CT	GA	GA	GA	CT	
50	0	+	62,0	B	CC	GG	GG	GG	CC	
48	0	+	68,1	B						
52,5	0	+	65,0	B	TT	AG	AG	AG	TC	
	0	+		A						
44	0	+	310,1	B	CT	GG	GG	GG	CC	
42	0	+	63,6	B	TT	AG	AG	AG	TC	
55	0	+	61,3	B	CT	GA	GA	GA	CT	
55	0	+	65,1	B	TT	AG	AG	AG	TC	
52	0	+	68,2	B						
58	0	+	59,7	B	CC	GG	GG	GG	CC	
48	0	+	66,6	B						
48	0	+	58,0	A	CT	GA	GA	GA	CT	
54	0	+	89,0	B						
48	0	+	65,0	C						
63	0	+	58,0	B	CT	GG	GG	GG	CC	
	0	+		A						

43	0	+	96,0	B					
44	0	+	63,4	B					
71	0	+	64	A					
52	0	+	59,2	D	CT	GA	GA	GA	CT
49	0	+	66	B					
48	0	+	55	B					
50	0	+	60	C					
39	0	+	58	B	CT	GA	GA	GA	CT
50	0	+	65	D					
58			58	B					
50	A	+	56,1	B					
47	A	+	68,3	C					
	0	+	62,4	A					
43	0	+	60,1	B	CT	GA	GA	GA	CT
49	0	+	56	B					
47	0	+	56	C					
52	0	+	60,3	C	TT	AG	AG	AG	TC
50	0	+	58,3	D					
49	0	+	58,1	B					
48	0	+	68,1	B					
52	0	+	59	D					
49	0	+	58	B					
52	0	+	65,1	B					
40	0	+	62,3	B	CT	GA	GA	GA	CT
53	0	+	58	B	CT	GG	GG	GG	CC
49	0	+	67	B					
54	0	+	59	B	CC	GG	GG	GG	CC
47	0	+	56,6	B					
45	0	+	66	B	CT	GG	GG	GG	CC

				A						
55	0	+	68	C	CT	GG	GG	GG	GG	CC
48	0	+	58	?						
68	0	+	64	B	TT	AA	AA	AA	AA	TT
45	0	+	58,7	B	TT	AA	AA	AA	AA	TT
53	0	+	61,3	C	TT	AA	AA	AA	AA	TT
44	0	+	58,4	B	TT	AG	AG	AG	AG	TC
52	0	+	59,2	B						
54	0	+	59,6	?						
50	0	+	58,3	C						
49	0	+	61,3	B						
49	0	+	65,4	D						
47	0	+	69,6	A						
53	0	+	62,5	B						
60	B	+	63,6	B	CT	GA	GA	GA	GA	CT
50	0	+	58,7	B						
49	0	+	65,3	C						
50	B	+	60,1	B						
50	0	+	61,2	?	TT	AA	AA	AA	AA	TT
52	0	+	62	B	TT	AG	AG	AG	AG	TC
58	0	+	63	B						
51	0	+	61,6	B						
48	0	+	64	C						
49	0	+	64,2	B						
49	0	+	63,2	B						
52	0	+	66	B	TT	AG	AG	AG	AG	TC
50	0	+	67,4	C						
49	0	+	62	B						
48	0	+	69,3	A						

43	0	+	68,1	B						
49	0	+	71,4	C						
49	B	+	64	B						
51	0	+	62	B						
48	0	+	62	D						
53	0	+	63	B						
60	0	+	63,9	B						
50	0	+	63	B						
49	0	+	62	C						
58	0	+	61,4	B						
50	0	+	64,6	B						
51	0	+	64,7	B						
46	0	+	60	B						
46	0	+	66,9	B						
49	0	+	61,4	B						
45	0	+	64,2	B						
49	0	+	62	B						
50	0	+	65,6	B						
50	0	+	62,3	C						
55	0	+	60	B						
50	0	+	63,2	B						
49	A	+	64,2	A						
50	0	+	63	C						
50	0	+	61	B						
54	0	+	63,7	B						
50	0	+	64	B						
54	0	+	64	D						
52	0	+	62,8	B						

Appendice 3_ Site Models

Per ogni gene in parentesi è indicata la lunghezza della sequenza allineata

Specie	<i>EDN1 (663bp)</i>	<i>EGLN1 (1332bp)</i>	<i>EGLN2 (1329bp)</i>	<i>EGLN3 (723bp)</i>
Bos grunniens				
Bos taurus	gi 74354035	ENSBTAG00000010500		gi 154426061
Callithrix jacchus	ENSCJAG00000005130	ENSCJAG00000018333		ENSCJAG00000005424
Canis lupus familiaris	gi 38603639		ENSCAFG00000005079	
Eospalax baileyi				
Eospalax cansus				
Equus caballus				ENSECAG00000005086
Felis catus	gi 56805590			
Gallus gallus				ENSGALG00000010001
Gorilla gorilla	ENSGGOG00000013300	ENSGGOG00000024267	ENSGGOG00000014955	ENSGGOG00000014900
Homo sapiens	gi 16307263	gi 14547145	gi 14547147	gi 14547149
Loxodonta africana	ENSLAFG00000007305			
Macaca mulatta	ENSMMUG00000007263	ENSMMUG00000005247	ENSMMUG00000001197	ENSMMUG00000000040
Monodelphis domestica				
Mus musculus	gi 1536791		gi 14547240	gi 14547242
Oryctolagus cuniculus	gi 1556	ENSOCUG00000013221	ENSOCUG00000023222	ENSOCUG00000008560
Ovis aries	gi 12239615			
Pan troglodytes	ENSPTRG00000017730		ENSPTRG00000011010	
Pongo abelii			gi 55732997:273-599	
Pongo pygmaeus	ENSPPYG00000016237	ENSPPYG00000000114		
Rattus norvegicus	gi 204067		gi 51858630	gi 9507122
Spalax carmeli				
Spalax ehrenbergi				
Spalax galili				
Spalax golani				
Spalax judaei				
Spermophilus tridecemlineatus			ENSSTOG00000004825	
Sus scrofa	gi 1945		ENSSSCG00000002995	ENSSSCG00000001963
Tursiops truncatus			ENSTTRG00000007185	
Vicugna pacos				
Tot. per gene	15	7	12	11

<i>HIF2a (EPAS1) (2649bp)</i>	<i>EPO (591bp)</i>	<i>HIF1a (2559bp)</i>	<i>NOS3 (3711bp)</i>	<i>VEGFA (753bp)</i>	<i>VHL (720bp)</i>
gi 148745434	gi 125976018	gi 111143452			
gi 296223953	gi 31343281	gi 117935054	gi 31341042	gi 27806356	gi 158517958
gi 73970125	ENSCJAT00000029305	gi 73963114	gi 50978891	gi 160358859	gi 55976934
		gi 78058355			
	gi 27807633	gi 148251121	gi 194210133	gi 126352555	
	gi 392793			gi 57619050	ENSFCAG00000011238
		gi 45383549		gi 160358851	
gi 262527236	gi 62240996	gi 194473733	gi 48762674	gi 284172458	ENSGGOG00000010456
	ENSLAFG00000027108		ENSLAFG00000007006	ENSLAFG00000032226	gi 2282063
	gi 342095	gi 109083867			
gi 126304465		ENSMODG00000008760		ENSMODG00000018950	
gi 1805269	gi 113931667	gi 226061947	gi 161086987	gi 118131137	
			gi 130502159		
	gi 395049		gi 194018577	gi 3228692	
gi 114577240	gi 114615072	gi 197098557			gi 197102443
gi 297667705					
gi 1283120	gi 8393315	gi 13242248	gi 46409655	gi 6606563	
	gi 48927338			gi 5919212	
	gi 48927342	gi 48927334			
	gi 48927336	gi 51921790			
	gi 48927340	gi 178056662	gi 47523659	gi 47522979	ENSSTOT00000013955
gi 148228021	gi 55742714	ENSTTRG00000012558			ENSSCG00000011567
	ENSTTRG00000011020	ENSVPAT00000002395			ENSTTRT00000007560
10	18	17	10	13	9

Tot. per sequenza

2
9
5
7
1
1
4
4
3
5
10
4
6
3
8
5
4
4
3
3
8
1
1
1
1
2
3
9
4
1

122

Appendice 4_Branch Models

	<i>EPO</i>	<i>HIF1a</i>	Tot. per specie
Bos grunniens	gi 125976018	gi 111143452	2
Bos taurus	gi 31343281	gi 117935054	2
Canis familiaris	gi 45826470	gi 73963114	2
Dipodomys ordii	ENSDORG00000014792	ENSDORG00000004304	2
Eospalax baileyi		gi 78058355	1
Eospalax cansus		gi 148251121	1
Equus caballus	gi 27807633	gi 194225073	2
Homo sapiens	gi 62240996	gi 194473733	2
Macaca mulatta	gi 342095	gi 109083867	2
Microtus oeconomus	gi 109894632	gi 63028438	2
Mus musculus	gi 113931667	gi 226061947	2
Oryctolagus cuniculus	gi 11527284	gi 130486011	2
Pantholops hodgsonii		gi 62632234	1
Rattus norvegicus	gi 8393315	gi 13242248	2
Spalax carmeli	gi 48927338		1
Spalax galili	gi 48927342		1
Spalax golani	gi 48927336		1
Spalax judaei	gi 48927340	gi 48927334	2
Sus scrofa	gi 55742714	gi 178056662	2
Tursiops truncatus	ENSTTRG00000011020	ENSTTRG00000012558	2
Vicugna pacos		ENSVPAT00000002395	1
Tot. per gene	17	18	35

Le specie adattate ad ambienti ipossici sono indicate in grassetto