

CARATTERIZZAZIONE GENETICA DEL CAVALLO NORICO ALLEVATO IN ITALIA

D'Agaro E.¹, Cimolino M.¹, Costa F.²

¹ DIPARTIMENTO DI SCIENZE - Università degli Studi di Udine

²Medico Veterinario, Libero Professionista

Riassunto

Il presente lavoro si propone di fornire un primo contributo alla conoscenza del grado di variabilità genetica del cavallo Norico allevato in Italia mediante la tipizzazione del gene D-loop (regione di controllo) mitocondriale della lunghezza di circa 478 pb. L'allineamento delle sequenze nucleotidiche ha evidenziato la presenza nella popolazione di 272 siti polimorfici e 7 aplotipi. La topologia dell'albero filogenetico della popolazione del cavallo norico del Friuli Venezia Giulia, Veneto e Alto adige è risultata congrua con la distribuzione geografica e simile per i diversi metodi di calcolo utilizzati.

Abstract

Genetic characterization of Pinzgauer horse reared in Italy - *In this study, a D-loop mitochondrial DNA (mtDNA) marker, was used to assess the genetic diversity in the Pinzgauer populations reared in the Friuli Venezia Giulia, Veneto and South Tirol regions. The aligned sequence consisted of about 478 bp. We found 272 polymorphic sites (56%) and 93 parsimony informative sites and seven haplotypes. The nucleotide frequencies were: A, 28.97; C, 29.39; G, 15.48 and T/U, 26.15%; average pairwise distance: 0.8454 and transition/transversion ratio: 0.52. Phylogenetic inference was performed using the Neighbour-Joining (NJ), the maximum likelihood (ML) (PhyML and RaxML) and MrBayes (MB) methods. Phylogenetic analysis resulted in a congruent tree topology for all the applied methods, in agreement with the population geographical distribution. A notable separation from haplotypes imported from Austria, the out-group and the other groups were observed. Information obtained, in the present study, should be used in the future to set up conservation programmes.*

Introduzione

Studio della biodiversità

La biodiversità indica una misura della varietà di specie animali e vegetali in un dato ecosistema ed è il risultato di lunghi processi evolutivi. L'evoluzione è il meccanismo che da oltre tre miliardi di anni permette alla vita di adattarsi al variare delle condizioni sulla terra. La diversità della vita sulla terra è costituita dall'insieme degli esseri viventi che popolano il pianeta. La biodiversità è intesa non solo come il risultato dei processi evolutivi, ma anche come il serbatoio da cui attinge l'evoluzione per attuare tutte le modificazioni genetiche e morfologiche che originano nuove specie viventi. La biodiversità si può considerare almeno in tre livelli diversi:

- a livello di geni in una specie;
- a livello di specie;
- a livello di ecosistemi.

Le caratteristiche morfologiche, ovvero tutte le caratteristiche visibili degli esseri viventi come ad esempio il colore degli occhi, il colore del pelo del mantello dei cavalli, sono esempi della varietà che esiste a livello di geni nell'ambito di ogni singola specie. La varietà di razze di cavallo sono esempi della biodiversità a livello di specie. Infine, la varietà di ambienti in una determinata area naturale e l'espressione della biodiversità a livello di ecosistema. Nell'ambito di una data comunità biologica è necessaria un'uniforme e approfondita conoscenza dei dati di base e la disponibilità di dati recenti. Elemento centrale nello studio della biodiversità animale, è una corretta pianificazione dell'uso del territorio che preveda la creazione di un'adeguata rete di aree protette destinate alla tutela integrale, allo sfruttamento eco-compatibile e allo sviluppo controllato delle specie animali a rischio di estinzione. La biodiversità esprime la variabilità genetica in un dato ambiente e viene studiata attraverso l'analisi dei loci polimorfici. Questi sono loci che variano considerevolmente tra gli individui. Se un locus ha due o più alleli, le cui frequenze superano il 5%, il locus viene definito polimorfico. Il locus polimorfico viene spesso definito un polimorfismo. Per procedere all'analisi dei polimorfismi genetici, si estrae il DNA dall'organismo target con metodi specifici di estrazione, poi, si amplifica il frammento o i frammenti di interesse mediante la PCR (Polymerase Chain Reaction) e successivamente si analizzano gli ampliconi (i prodotti risultanti dalla reazione di PCR) per identificare dei caratteristici polimorfismi genetici. Verranno, a seguito, descritti i principali passaggi per l'analisi dei polimorfismi (estrazione, amplificazione e analisi del DNA). La genetica di popolazione ha lo scopo di descrivere la struttura della variabilità genetica e di determinare le forze evolutive che l'hanno formata. Questo approccio è fondamentale per una corretta comprensione dell'evoluzione, definibile come l'insieme dei cambiamenti progressivi che subiscono le frequenze alleliche nelle popolazioni. Gli studi di genetica di popolazione hanno due principali finalità: misurare la quantità di variazione genetica esistente nelle popolazioni e spiegare come questa viene originata e mantenuta, nonché quale sia la sua importanza dal punto di vista ecologico ed evolutivo. Il dato di base di questi studi è fornito dalle frequenze alleliche di geni polimorfici. In organismi diploidi, ogni individuo ha due copie di ogni gene o locus, e può quindi presentare tre differenti combinazioni alleliche, o genotipi. Gli individui che portano due copie dello stesso allele sono detti omozigoti, mentre sono detti eterozigoti gli individui che presentano due alleli diversi. Dalle frequenze genotipiche di una popolazione, è possibile ottenere le frequenze alleliche, che costituiscono la più semplice misura della variabilità presente. Una misura più utile è data dalla frequenza totale degli eterozigoti nella popolazione, o eterozigosità (h), per ogni locus. È possibile calcolare anche l'eterozigosità media tra tutti i loci (H). Se il calcolo è ottenuto dalle frequenze genotipiche osservate, si ottiene l'eterozigosità media osservata, mentre se si utilizzano le frequenze alleliche attese in una popolazione panmittica secondo il teorema di Hardy-Weinberg), questa viene definita eterozigosità attesa. La deviazione dei dati osservati da quelli attesi indica se le assunzioni alla base della stima non sono rispettate, suggerendo la possibilità di strutture interne alla popolazione, di migrazioni o di effetti dovuti alla pressione selettiva. La struttura della variabilità genetica delle popolazioni

viene, infatti, determinata da diversi fattori, quali le suddivisioni entro e tra popolazioni, l'inbreeding e il flusso genico. Per valutare questi fattori, è importante stabilire il grado di deviazione dell'eterozigosita osservata da quella attesa, e questo viene solitamente effettuato utilizzando la statistica F.

Marcatori molecolari

La variabilità genetica è determinata da ogni variazione dei nucleotidi nel genoma degli organismi. Lo studio della variabilità genetica è facilitato dall'utilizzo di marcatori, che rappresentano caratteristiche ereditabili e polimorfiche a livello di specie, popolazioni ed individui. Un marcatore molecolare è un qualsiasi locus le cui varianti alleliche possono essere identificate facilmente analizzando direttamente il DNA. Un buon marcatore molecolare deve essere polimorfico, codominante, facilmente individuabile, ripetibile, neutrale rispetto alla selezione e deve possedere un meccanismo di trasmissione semplice (mendeliano o uniparentale). I marcatori morfologici, che considerano le caratteristiche fenotipiche, e numerosi marcatori molecolari (allozimi) sono influenzati enormemente dall'ambiente, mentre i gruppi sanguigni e i polimorfismi del DNA determinati da mutazioni non adattative sono in genere marcatori neutri. Gli allozimi sono le varianti alleliche degli enzimi codificati da geni strutturali che differiscono in base al numero di sostituzioni aminoacidiche. La mutazione di un aminoacido porta alla variazione della carica della proteina e tale differenza può essere visualizzata tramite corsa su gel elettroforetico. Gli allozimi sono il primo esempio di marcatore molecolare e sono stati ampiamente utilizzati in studi di genetica di popolazione e sulla biodiversità. Essendo marcatori poco abbondanti e poco polimorfici, sono stati sostituiti quasi completamente da marcatori basati sul DNA, che hanno come caratteristiche quelle di elevata variabilità (alto numero di loci polimorfici o di alleli per locus) e di facile identificazione. Le variazioni della sequenza genomica, che possono essere dovute a cambiamenti di singole basi, numero di ripetizioni di corte sequenze di- o tri-nucleotidiche, inserzioni, delezioni o duplicazioni, sono identificate come polimorfismi e determinano l'esistenza di diversi alleli. La tecnica di biologia molecolare attualmente più utilizzata per l'individuazione dei differenti alleli si basa sulla determinazione degli SNPs (Single Nucleotide Polymorphism) basata sull'utilizzo della tecnica PCR e del successivo sequenziamento dell'amplicone.

Marcatori mitocondriali per indagini biogeografiche

Il genoma mitocondriale, presente in tutti i metazoi, è costituito da una molecola di DNA circolare lunga in media 15-20 kp (5µm di lunghezza), localizzata vicino alle creste mitocondriali. La molecola di mtDNA è composta di 37 geni, 13 dei quali codificano per proteine coinvolte nei meccanismi respiratori della cellula, 22 codificano per tRNA, 2 codificano per RNA ribosomali. Il genoma mitocondriale ha una struttura compatta, senza regioni introniche e regioni spaziatrici fra i diversi geni. Alcune caratteristiche del mtDNA, lo rendono particolarmente adatto, per i diversi tipi di indagine genetica. La molecola è di modeste dimensioni ed è presente con un elevato numero di copie in ogni cellula. Il genoma mitocondriale non si trasmette secondo un modello ereditario mendeliano.

no, ma attraverso il meccanismo dell'eredità materna (solo la madre trasmette il proprio mtDNA alla prole, senza eventi di ricombinazione). Nel mtDNA sono presenti delle regioni di DNA altamente conservate, adatte, quindi, per progettare oligonucleotidi da impiegare come primers universali in reazioni di polimerizzazione a catena del DNA (PCR), mentre altre regioni mostrano un elevato tasso di mutazione. Il tasso di mutazione è, in genere, più elevato rispetto a quello del genoma nucleare e la maggior parte delle sostituzioni che si riscontrano tra i DNA mitocondriali sono mutazioni puntiformi, con una forte preponderanza di transizioni (sostituzioni purina → purina o pirimidina → pirimidina) rispetto a trasversioni (sostituzioni purina → pirimidina o viceversa). Quest'alto tasso di mutazione è dovuto ad un'elevata frequenza di errori durante la replicazione, causati sia dall'inefficienza dei meccanismi di riparazione del DNA, sia dall'ambiente del citoplasma mitocondriale, fortemente ossidante per la presenza di radicali liberi e superossidi. Inoltre, la pressione selettiva nei mitocondri non è efficace nei mitocondri come nel genoma nucleare: cambiamenti nelle proteine, nei tRNA e nei rRNA codificati dalle sequenze del mtDNA, hanno un effetto minore sul fitness dell'individuo. Poiché, le mutazioni sono accumulate a un ritmo che può essere considerato costante, il numero di mutazioni che differenziano il DNA mitocondriale di un individuo da quello dei suoi antenati può essere utilizzato come un orologio molecolare per calcolare il tempo che li separa, fornendo uno strumento utile ma molto delicato per la ricostruzione della filogenesi degli organismi. Negli studi di popolazioni, le analisi con i marcatori mitocondriali forniscono informazioni relative alla sola linea femminile.

La bioinformatica

La genomica moderna non avrebbe avuto lo sviluppo attuale senza l'ausilio dei computer. I programmi (software) sono essenziali per l'assemblaggio, delle letture contig e dei genomi. Inoltre, i mezzi di memorizzazione ad alta capacità sono essenziali per creare e mantenere gli archivi dei dati. La distribuzione delle informazioni e la disponibilità degli strumenti analitici, richiedono le funzioni fornite dalle reti di computer e il World Wide Web. L'informatica è una disciplina teorica che fornisce un background intellettuale sia nella progettazione che nell'implementazione di metodi per la risoluzione di problemi.

Analisi degli algoritmi

Per la ricerca di sequenze simili si deve comparare la sequenza sonda con tutte le sequenze presenti nel database. Un campo specifico dell'informatica, detto gergalmente "stringology" si occupa dello sviluppo di algoritmi efficienti per la risoluzione di questo tipo di problemi e sull'analisi delle loro prestazioni attese. Gli algoritmi specificano i termini e i programmi, poi, li implementano. L'efficacia di un programma dipende dall'algoritmo su cui si basa e dalla capacità del programmatore.

Allineamento di sequenze

Date due o più sequenze, si desidera:

- misurare la loro similarità;

- comprendere le corrispondenze reciproche fra i residui;
- osservare pattern di conservazione e variabilità;
- dedurre affinità evolutive.

Un'importante applicazione dell'allineamento di sequenze è l'annotazione di geni, attraverso l'identificazione di omologhi e l'assegnazione della struttura e funzione. L'allineamento di sequenze consente l'identificazione di corrispondenze residuo-residuo. Per decidere quale sia il migliore degli allineamenti, è necessario un metodo per esaminare sistematicamente tutti gli allineamenti possibili. Si deve calcolare un punteggio (score) che rispecchi la qualità di ogni allineamento possibile e identificare un allineamento con il punteggio ottimale. L'allineamento ottimale può non essere unico: molti differenti allineamenti possono dare lo stesso punteggio. Inoltre, anche variazioni minori dello schema di assegnazione del punteggio possono variare l'ordinamento gerarchico (ranking) degli allineamenti, facendo sì che emerga come allineamento migliore un allineamento diverso.

Diagramma a punti

Il diagramma a punti (dot plot) è una rappresentazione semplice che offre uno sguardo d'insieme sulla similarità tra coppie di sequenze. Meno evidente è la sua stretta relazione con gli allineamenti. Il diagramma a punti è una tabella o una matrice. Le righe corrispondono ai residui di una sequenza e le colonne alle righe dell'altra sequenza. Nella sua forma più semplice, le posizioni nel diagramma a punti sono lasciate vuote se i residui sono diversi e vengono riempite se i residui sono in corrispondenza. Tratti di residui simili si presentano come diagonali nella direzione alto a sinistra - basso a destra. I diagrammi a punti consentono di tracciare i grafici della relazione tra due sequenze. Per le applicazioni di biologia molecolare, vengono assegnate pesi variabili a differenti operazioni di edit. Per esempio, le sostituzioni aminoacidiche tendono a essere conservative: la sostituzione di un amminoacido con un altro di dimensioni o di proprietà fisico-chimiche simili è più probabile della sua sostituzione con un altro amminoacido con proprietà dissimili. Allo stesso modo, la delezione di più basi contigue è più probabile della delezione indipendente dello stesso numero di basi. Un programma per computer è in grado di assegnare un punteggio a ogni percorso attraverso il diagramma a punti sommando i punteggi dei singoli passi. Per ciascuna sostituzione, addiziona il punteggio della mutazione, a seconda della coppia di residui implicata. Per le mosse orizzontali e verticali, aggiunge un'appropriata penalità per i gap.

Sistemi di assegnazione del punteggio

Un sistema di assegnazione del punteggio (scoring) deve tenere conto delle sostituzioni di residui e delle inserzioni o delezioni. Le delezioni, ossia i gap in una sequenza, avranno punteggi dipendenti dalla loro lunghezza. Nel caso delle sequenze degli acidi nucleici, si usa comunemente un semplice schema per le sostituzioni, + 1 per un match, - 1 per un mismatch, o uno schema più complicato basato sulla frequenza delle mutazioni per transizione più alta della frequenza delle mutazioni per transversione. Nel caso delle proteine, sono stati

proposti vari sistemi di scoring. Si potrebbero raggruppare gli amminoacidi in classi di tipo fisico-chimico simile e assegnare un punteggio + 1 per un match entro una classe di residui e un punteggio -1 per i residui in differenti classi. Si potrebbe tentare di ideare un metodo più preciso, per attribuire un punteggio alle sostituzioni in base ad una combinazione di proprietà degli amminoacidi. Oppure, si potrebbe tentare di lasciare che siano le proteine a fornire un appropriato sistema di scoring. Dayoff è stata la prima a farlo, raccogliendo statistiche sulle frequenze di sostituzione nelle sequenze proteiche allora conosciute. I suoi risultati sono stati usati per molti anni per attribuire punteggi agli allineamenti. Questo metodo è stato poi sostituito da un altro che si basa su matrici più complesse.

Matrici BLOSUM

Le matrici BLOSUM sono state sviluppate dai biologi molecolari statunitensi Henikoff e Henikoff per lo scoring delle sostituzioni nei confronti tra sequenze amminoacidiche. Le matrici BLOSUM si basano sul database BLOCKS di sequenze proteiche allineate, da cui il nome: BLOcks SUBstitution Matrix. Basandosi su regioni di proteine strettamente correlate allineabili senza gap, Henikoff ha calcolato il rapporto tra il numero di coppie osservate di amminoacidi in ogni posizione e il numero di coppie attese in base alle frequenze di amminoacidi complessive. Per evitare di attribuire un peso eccessivo alle sequenze strettamente correlate, Henikoff ha sostituito gruppi di proteine che avevano identità di sequenza superiori a una certa soglia, con una singola proteina rappresentata oppure con una media ponderata. La soglia del 62% produce la matrice di sostituzioni BLOSUM62 di uso comune. Il calcolo è il seguente. La entry 2 della matrice corrisponde al valore effettivo 0,2 in virtù del fattore di scala. Il valore 0,2 è il logaritmo in base 10 del valore atteso relativo della mutazione. Dato che $\log_{10}(1,6) = 0,2$, il valore atteso è + 1,6. La probabilità di due eventi mutualmente indipendenti è il prodotto delle loro probabilità. Usando i logaritmi si ottengono punteggi che possono essere addizionati invece che moltiplicati, è quindi conveniente ai fini dell'elaborazione.

Attribuzione di punteggi a inserzioni e delezioni, o ponderazione dei gap

Per ottenere un sistema completo di scoring per gli allineamenti è necessario, oltre alla matrice di sostituzione, un metodo per attribuire un punteggio ai gap. Per l'allineamento delle sequenze di DNA, il pacchetto di software di allineamento ampiamente usato CLUSTALW raccomanda di usare la matrice di identità (o matrice identica) per la sostituzione (+ 1 per un match, 0 per un mismatch), e penalità di gap 10 per l'inizio di un gap e penalità di gap 0,1 per l'estensione di un gap di un residuo. Per l'allineamento delle sequenze proteiche, si raccomanda di usare la matrice BLOSUM62 per le sostituzioni, con penalità di gap 11 per l'inizio di un gap e 1 per l'estensione di un gap di un residuo.

Metodi approssimati per lo screening rapido di un database

È pratica routinaria, sottoporre a screening i geni di un nuovo genoma confrontandoli con un database, per trovare similarità con altre sequenze. I data-

base sono diventati così grandi che i programmi basati su allineamenti locali sono troppo lenti. I metodi approssimati sono capaci di identificare bene e rapidamente le affinità strette ma sono meno precisi rispetto ai metodi esatti nel cogliere affinità molto lontane. In pratica, danno una performance soddisfacente quando la sequenza sonda è abbastanza simile a una o più sequenze presenti in una banca dati. Un tipico approccio di approssimazione è il BLAST (Basic Local Alignment Search Tool). Una sequenza candidata è una sequenza presente nel database con una spaziatura equivalente nella sequenza sonda e nella sequenza candidata. Per una serie selezionata di sequenze candidate, vengono eseguiti i calcoli approssimati dell'allineamento ottimale, tenendo conto che i percorsi attraverso la matrice considerata siano limitati a bande intorno alle diagonali.

Costruzione di alberi filogenetici

Gli alberi filogenetici sono grafici bidimensionali in grado di visualizzare le relazioni evolutive che intercorrono tra diversi organismi, sequenze o geni. In linea generale, tanto più c'è similitudine genetica fra due elementi all'interno dell'albero, tanto maggiore sarà la vicinanza fra di essi. I nodi rappresentano un evento di separazione, o la creazione di due entità distinte a partire da un antenato comune. I metodi utilizzati per la costruzioni di alberi filogenetici possono basarsi su singoli caratteri (*Maximum Parsimony* e *Maximum Likelihood*) oppure sulle distanze fra le sequenze (*Minimum Evolution*). I sistemi di *Minimum Evolution* comunemente utilizzati sono: *Neighbor-Joining* e *UPGMA*. *Maximum Parsimony* è un sistema che ricerca l'albero che spiega le sequenze osservate con il numero minimo di sostituzioni. L'algoritmo prevede di trovare tutti gli alberi possibili, di valutare il "costo" di un albero in termini di mutazioni e di scegliere, infine quello con il costo inferiore. *Maximum Likelihood* è un sistema che valuta la probabilità che i dati di sequenza osservati siano stati prodotti da un determinato percorso evolutivo e filogenetico. L'algoritmo prevede un modello filogenetico, di ricercare tutti gli alberi che è possibile costruire in funzione di questo modello, di valutare la probabilità che gli allineamenti ottenuti derivino dalle relazioni rappresentate da un dato albero e di scegliere, infine l'albero con maggiore verosimiglianza. *MrBayes* utilizza per il calcolo delle distanze un metodo introdotto recentemente di inferenza bayesiana, che sta dando ottimi risultati in termini di qualità degli alberi generati. *Neighbor-Joining* è un sistema che per costruire un albero filogenetico non considera semplicemente la distanza tra le coppie, ma la mette in rapporto alla distanza media fra tutti gli altri punti. Esso produce alberi in cui la lunghezza dei bracci esprime la distanza genetica fra i taxa. La topologia dell'albero risulta corretta se l'evoluzione delle sequenze non segue il modello dell'orologio molecolare, cioè se non esiste una relazione lineare tra divergenza di sequenza ed il tempo evolutivo. *UPGMA* (Unweighted Pair Group Method using arithmetic Average) è un sistema che raggruppa le sequenze partendo dalle più simili ed aggiungendo via via un nodo all'albero, procedendo dalle "foglie" verso la radice". Esso produce alberi in cui la lunghezza dei bracci esprime il tempo di separazione tra i taxa. Gli alberi filo-

genetici prodotti risultano senza radice (*unrooted*). Essi possono comunque essere orientati inserendo una sequenza più lontana (*outgroup*) della quale si conosca la relazione con i taxa studiati. La stima statistica dell'affidabilità della costruzione di un albero viene individuata dal *bootstrap*. Il *bootstrap* viene indicato per ogni nodo e corrisponde al numero di volte che il nodo stesso è stato ottenuto considerando un certo numero di campionamenti casuali partendo dal database originale.

Evoluzione storica del cavallo Norico

Il cavallo Pinzgauer, noto anche come il cavallo Norico, è una razza austriaca da tiro moderatamente pesante. Il cavallo Norico è considerato autoctono della regione alpina dell'Europa centrale e si ritiene abbia avuto origine nella regione di Salisburgo a partire dall'anno 1897. Questa regione, che si estendeva a sud nell'attuale Carinzia e Slovenia, divenne nota come la provincia romana "Noricum". Alla fine del XIX secolo, il nome del cavallo Pinzgauer è stato cambiato in Norico, in relazione all'atteggiamento "romanofilo" di quell'epoca. Il cavallo Norico ha svolto un ruolo importante per diversi secoli, nel trasporto di merci attraverso le Alpi. Tuttavia, negli ultimi decenni, i cavalli da tiro pesante hanno perso progressivamente importanza in relazione alla progressiva meccanizzazione dei sistemi agricoli e dei trasporti. Per questo motivo, la numerosità della popolazione del cavallo Norico allevato in Austria e Italia si è ridotta in modo allarmante alla fine degli anni settanta. Fortunatamente, il sostegno finanziario da parte del governo austriaco e di alcuni enti locali e la dedizione degli allevatori di cavallo Norico hanno contribuito a salvare questa razza dall'estinzione. Nella regione Friuli Venezia Giulia l'importazione e utilizzo di femmine e stalloni di cavallo Norico dall'Austria è iniziata solo recentemente mentre gli interscambi tra l'Austria e l'Alto Adige sono stati più frequenti e duraturi nei decenni passati. La popolazione di cavallo Norico, che è attualmente presente nel territorio del Friuli Venezia Giulia, prende origine principalmente da animali allevati localmente e/o recentemente incrociati con animali austriaci. In diverse ricerche e indagini genetiche condotte sui cavalli è stato utilizzato come marcatore genetico un gene della regione D-loop mitocondriale (Cozzi et al., 2004). Questo gene, che costituisce la parte più variabile del mtDNA in relazione a un tasso di sostituzione più elevato rispetto al resto del genoma mitocondriale è utilizzato come marcatore genetico di riferimento (barcoding) negli studi filogenetici del cavallo. Sono riportati a seguito i risultati preliminari della caratterizzazione genetica del cavallo Norico nella regione Friuli Venezia Giulia, Alto Adige e Belluno.

Materiali e Metodi

Nella presente indagine sono stati considerati 36 soggetti di cavallo Norico, distribuiti nelle province di Udine, Gorizia e Trieste in Friuli Venezia Giulia, Bolzano in Alto Adige e Belluno in Veneto. Gli animali, allevati in 15 allevamenti, sono stati tipizzati per il gene D-loop mitocondriale. Nella Tabella 1 vengono riportati i siti di campionamento nelle tre regioni.

Tabella 1 - Siti di campionamento e numero dei campioni del cavallo Norico.

N°	Località	N
1	Cividale	4
2	Strassoldo	2
3	S. Giorgio di Nogaro	2
4	Cormons	1
5	Faedis	1
6	Trieste	3
7	Bolzano	10
8	Versciaco	3
9	Prato alla Drava	3
10	Sesto	4
11	Belluno	3

Estrazione DNA

Il DNA è stato prelevato dalle cellule boccali mediante tampone Omni Swab (Whatman), da 28 cavalli. Il protocollo di estrazione del DNA era il seguente:

- prelievo di 1,2 – 2 mm di tampone (Whatman) contenente il campione e depositato in una provetta da 1,5 mL;
- aggiunta di 200 µl di FTA purification reagent (Whatman);
- incubazione a temperatura ambiente per 5 min agitando la provetta manualmente;
- rimozione del surnatante;
- aggiunta 200 µl di TE Buffer (10mM Tris – HCl, 0.1mM EDTA pH 8.0);
- incubazione a temperatura ambiente per 5 minuti agitando la provetta manualmente;
- rimozione del surnatante;
- lisi alcalina: aggiungere 35 µL di Buffer Alcalino (0.1N NaOH, 0.3mM EDTA pH13);
- incubazione a 95°C per 3 minuti (Thermomixer Comfort Eppendorf);
- aggiunta di 65 µL di soluzione neutralizzante (0.1M Tris – HCl pH 7.0) e agitare la provetta manualmente.

Il DNA estratto e normalizzato è stato amplificato in un volume di reazione di 25µL contenente: 2.5mM di MgCl₂ (Sigma, Milano), 0.2mM di dNTPs (AB, UK), 0.5µM di ogni primer (Sigma, Milano), 1U di Taq polimerasi (Sigma, Milano), 1x di PCR Buffer (Sigma, Milano) 100 ng di DNA. La sequenza dei primer utilizzati era la seguente: NoricoF 5'-CGCACATTACCCTGGTCTTG- 3', e NoricoR 5'-GA-ACCAGATGCCAGGTATAG-3'. Per l'amplificazione è stato utilizzato il seguente protocollo: 95°C per 5 minuti, 35 cicli a 94°C per 40 secondi, 52°C per 45 secondi, 72°C per 45 secondi, seguiti infine da una fase di allungamento a 72°C per 10 minuti (Thermo Hybaid, Ashford, UK). Successivamente, i prodotti della prima reazione sono stati utilizzati come template per un'ulteriore reazione di

amplificazione. Infine, i prodotti ottenuti sono stati purificati utilizzando il kit Quiagen (QuiAquick PCR Purification Kit) e sequenziati utilizzando il metodo Big Dye Terminator (Applied Biosystem) con un sequenziatore automatico ABI 3770. Per tutti i campioni sono state ottenute le sequenze forward e reverse. Allineamenti multipli delle sequenze sono stati analizzati mediante il software DNASP versione 4.0 (Rozas et al., 2003) e mega 3.1. La diversità aplo-tipica (H_d) è stata calcolata usando le equazioni suggerite da Nei (1987). La diversità nucleotidica è stata stimata a partire dal numero S di siti polimorfici (Watterson 1975; Nei, 1987). I calcoli sono basati su tutti i siti segreganti (silenti e non sinonimi). Il test statistico di neutralità D , D^* (Tajima, 1989) e F^* (Fu and Li, 1993) sono stati calcolati usando 10000 simulazioni che hanno verificato le ipotesi che nel gene, le mutazioni selezionate fossero neutre (Kimura, 1983). L'inferenza filogenetica è stata eseguita utilizzando le analisi Neighbor-Joining (NJ), maximum likelihood (ML) (PhyML and RaxML) e Mr Bayes (MB) mediante il software TOPALI V2.5. Il modello ottimale dell'evoluzione nucleotidica per le analisi NJ e ML, è stato determinato usando il model test implementato nel programma. Il modello di selezione utilizza il metodo ML (in parallelo sui cluster o macchine multi core). I risultati sono ordinati in ordine crescente in base al punteggio BIC ($BIC = -2 \ln l(\log \text{somiglianza}) + df(\text{gradi di libertà}) + \log(n)$ (quantità di campione)). Il modello di selezione è stato testato sul tasso di eterogeneità (proporzione di siti invariati e distribuzione gamma). Il modello di selezione per Mrbayes è stato testato come sottoinsieme di modelli superiori.

Risultati e discussione

Nelle diverse popolazioni di cavallo norico analizzate, è stato individuato un moderato polimorfismo della regione ipervariabile del D-loop mitocondriale. L'allineamento con la tecnica pairwise ha evidenziato un'elevata similarità (97%) con le sequenze depositate nei database internazionali (GeneBank X79547, Xu e Arnason, 1994) e 0 gaps. Una descrizione dettagliata dei risultati ottenuti nella caratterizzazione genetica delle popolazioni di cavallo norico allevata nel Friuli Venezia Giulia, Veneto e Alto Adige è riportata nella Tabella 2.

Tabella 2 - Analisi statistica delle sequenze allineate per gli individui di cavallo noricoIndici di diversità degli aplotipi

Numero di aplotipi, h: 7

Indici di diversità degli aplotipi Hd: 1,000

Deviazione standard: 0,177

Indice di diversità nucleotidica

Indice di diversità nucleotidica, Pi: 0,06147

Deviazione standard di Pi: 0.001536

Indice di diversità nucleotidica (Jukes e Cantor), Pi(JC): 2,2267

Theta (per sito) da Eta: 0,7063

Theta (per sito) da S, Theta-W: 0,4975

Varianza di Theta : 0,0717

Deviazione standard di Theta: 0,2678

Modello per un numero limitato di loci

Theta (per sito) da Pi: 3,4078

Theta (per sito) da S: 2,1071

Theta (per sito) da Eta: 4.0982

Numero medio di differenze nucleotidiche, k: 293,83

Varianza stocastica, Vst(k): 13527,78

Varianza di k , Vs(k): 12375,98

Varianza totale di k , V(k): 25903,76

Test di neutralità

D di Tajima: -1.3632

Significatività statistica P <0.001

D* di Fu e Li: -0.3932

Significatività statistica, P > 0.10

F* di Fu e Li: -0,6150

Significatività statistica, P > 0.10

S di Strobeck : 3.885 (Probabilità che NHap <= 4)

Probabilità che [NHap = 4]: 0,98

La lunghezza della sequenza era pari a circa 478 pb. Sono stati identificati 272 siti polimorfici (56%) e 93 siti informativi per la parsimonia. La frequenza media dei nucleotidi nelle sequenze allineate era pari a: A, 28,97; C, 29,39; G, 15,48 e T/U, 26,15%; distanza media pairwise: 0,84; rapporto transizione/transversione: 0,52. L'indice di diversità degli aplotipi (Hd) e nucleotidica (Pi (π)) sono risultati pari rispettivamente a: $1 \pm 0,177$ (media \pm DS) e $0,06147 \pm 0,0015$. L'indice di diversità nucleotidica (il numero medio di differenze nucleotidiche per sito tra due sequenze) calcolato second il metodo di Jukes e Cantor è risultato pari a: 2,2267. Theta (θ) (per sito) è stato calcolato da Eta (numero totale mutazioni) o da S (il numero di siti polimorfici) per base. Theta (θ) = 4Nm

per un gene autosomico di un organismo diploide (N e m sono rispettivamente la numerosità effettiva e il tasso di mutazione per ciascun nucleotide per generazione). La varianza di questa stima dipende dal tasso di ricombinazione tra i siti. Queste varianze sono state calcolate per ciascun nucleotide: varianza (per nucleotide) = varianza (per sequenza di DNA) / m^*m dove m è il numero totale di nucleotidi considerati. Il livello di polimorfismo viene determinato mediante la stima di θ (Watterson 1975). Il confronto delle diverse regioni si ottiene considerando θ per ciascun sito. Nel presente studio, θ era pari a 0,4975, valore in linea con i dati della letteratura (Fu and Li, 1993). Questo parametro sintetizza l'effetto combinato della mutazione e della deriva genetica assumendo che la selezione non sia operativa. D , D^* e F^* hanno evidenziato dei valori negativi al test statistico di neutralità, ma solo per D questo effetto è risultato significativo. Il test D^* si basa sulle differenze tra il numero mutazioni puntiformi e il numero totale di mutazioni. Il test F^* si basa sulle differenze tra il numero di mutazioni puntiformi e il numero medio di differenze nucleotidiche tra paia di sequenze. Valori negativi per questi indici, indicano che è in atto una selezione negativa o l'eliminazione di un gene mutato e viceversa (Kreitman, 1983). È difficile, tuttavia, distinguere tra una popolazione in crescita o sottoposta a selezione se consideriamo solo la variabilità intra-specifica. Il confronto della variabilità intra-specifica e inter-specifica tra due specie simili può essere utilizzato per confermare l'ipotesi che i geni siano sottoposti a selezione (Kreitman, 1983; McDonald e Kreitman, 1991). Il test D di Tajima tende ad assumere valori negativi se ci sono molti geni rari (frequenze basse) e viceversa valori positivi con molti geni comuni (Tajima, 1989). Il miglior modello di sostituzione genica è risultato il modello H81 (Hasegawa *et al*, 1985). L'analisi filogenetica dei dati ha consentito di identificare sette aplotipi e due cluster principali per le popolazioni di cavallo norico allevate nel Friuli Venezia Giulia, Veneto e Alto adige. Gli alberi filogenetici, stimati con diversi metodi, Neighbor-Joining (NJ), MrBayes (MB) e ML (PhyML and RaxML), sono risultati simili. Nella Figura 1 viene riportato il dendrogramma ottenuto con il metodo MrBayes. I risultati dell'analisi statistica non hanno evidenziato differenze per i tre metodi e la topologia degli alberi filogenetici sono risultati congrui con la distribuzione geografica. La veridicità delle relazioni genetiche è stata valutata utilizzando l'analisi di bootstrap non parametrica, con 100 repliche per il modello ML e 1000 repliche per gli alberi filogenetici analizzati con MEGA 3.1. I valori di bootstrap uguali o maggiori di 60 sono da considerarsi ben supportati.

Fig. 1 – Albero MrBayes ottenuto mediante l'allineamento delle sequenze di cavallo norico e l'outgroup



Conclusioni

In conclusione, nel presente studio, l'analisi del polimorfismo al locus D-loop mitocondriale ha evidenziato una discreta variabilità nelle popolazioni di cavallo Norico allevate nelle regioni Friuli Venezia Giulia, Veneto e Alto Adige. Sono stati individuati complessivamente 7 aplotipi e due cluster principali.

Ringraziamenti

Si ringraziano l'Associazione Epona e l'Associazione Allevatori della provincia di Bolzano per la fattiva collaborazione.

Bibliografia

- Cozzi M.C., Strillacci, M.G., Valiati, P., Bighignoli, M., Cancedda, M., Zanotti, M, 2004. *Mitochondrial D-loop sequence variation among Italian horse breeds*. Genet. Sel. Evol., 36: 663–672.
- Fu Y. X., Li W. H., 1993. *Statistical Tests of Neutrality of Mutations*. Genetics, 133: 693-709.
- Hasegawa, M., Kishino, H. Yano, T., 1985. *Dating of the human-ape splitting by a molecular clock of mitochondrial DNA*. J. Mol. Evol., 22: 160–174.
- Kimura M. 1983. *The neutral theory of molecular evolution*. Cambridge Univ. Press, Cambridge, UK.
- Kreitman, M., 1983. *Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster*. Nature, 304: 412-417.
- McDonald J. H., Kreitman M., 1991. *Adaptive protein evolution at the Adh locus in Drosophila*. Nature, 351: 652-654.
- Nei M., 1987. *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.
- Rozas, J. C., Sanchez-Delbarrio X., Rozas R., 2003. *DnaSP, DNA polymorphism analyses by the coalescent and other methods*. Bioinformatics, 19: 2496-2497.
- Tajima F., 1989. *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism*. Genetics, 123: 585-595.
- Xu X, Arnason U. *The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region*. Gene, 148(2): 357-62.
- Watterson G.A., 1975. *On the number of segregating sites in genetical models without recombination*. Theor. Pop. Biol., 7: 256-276.