

Facoltà di Scienze
matematiche, fisiche e naturali

I GENI E L'AMBIENTE **(PRODOTTO DALL'UOMO)**

Luigi Luca Cavalli-Sforza

Laurea Honoris causa
in Ecologia

Parma, 30 novembre 2006

Lectio doctoralis di Luigi Luca Cavalli-Sforza
in occasione del conferimento
della Laurea Honoris causa in Ecologia

• Genomi ed Evoluzione	Pag.	5
• La fondamentale importanza della demografia per comprendere e misurare l'evoluzione.	Pag.	8
• La frequenza dei polimorfismi.	Pag.	10
• L'uomo come specie cosmopolita ed invasiva.	Pag.	15
• Il contributo della mutazione e della selezione al polimorfismo.	Pag.	21
• Colli di bottiglia demografici seguiti da crescita demografica.	Pag.	25
• L'effetto del fondatore in serie nell'espansione dell'uomo moderno.	Pag.	28
• La correlazione tra distanze geografiche e distanze genetiche.	Pag.	29
• Variazione genetica tra ed entro popolazioni ed il problema della razza.	Pag.	31
• Dovremo fare diagnosi e terapia su misura per ciascun genoma individuale?	Pag.	35
• Il progetto Human Genetic Diversity Project (HGDP) e il suo futuro.	Pag.	38
• Citazioni Bibliografiche.	Pag.	41

I GENI E L'AMBIENTE (PRODOTTO DALL'UOMO)

Luigi Luca Cavalli-Sforza

GENOMI ED EVOLUZIONE

Fino a tempi recentissimi lo studio di un intero genoma pur obiettivo affascinante pareva un sogno irrealizzabile. La situazione è cambiata e sono già quindici i genomi di specie completamente sequenziati. Il confronto di specie filogeneticamente distanti come un batterio ed un eucariote (tra cui l'uomo) sta rendendo possibile, con accuratezza crescente, la ricostruzione dei passi che hanno portato dai genomi ancestrali che sono andati persi a quelli che sopravvivono ai nostri giorni.

L'analisi dei genomi sta portando chiaro supporto alla supposizione per prima proposta da Lamarck, e molto diffusa all'inizio del diciannovesimo secolo, che tutti gli organismi viventi avevano un'origine da un antenato comune, anche se la definizione precisa della natura di questo organismo primigenio dovrà attendere ulteriori ricerche. La scienza ha dovuto combattere paure e scetticismo per tutta la sua esistenza, e non è quindi troppo sorprendente che alcuni non accettino l'evoluzione, nemmeno come un'ipotesi, e tentino di sostituirla con alternative

basate sui credi religiosi, anche in un momento in cui la genomica sta trasformando l'evoluzione in una delle teorie più potenti dell'umanità.

I dubbi e la discussione delle osservazioni e delle loro interpretazioni sono, naturalmente, lo stimolo di base per il progresso scientifico, ma bisogna ammettere che l'elevata proporzione di Americani che non accetta l'evidenza in favore dell'evoluzione è umiliante e preoccupante, specialmente se si tiene conto che sono cittadini del paese che è stato la culla di molta della scienza e della tecnologia più recente. Naturalmente siamo consapevoli che questo errore si basa sul fatto che, nella Genesi, è implicitamente stabilito che la vita sulla terra è esistita per non molto di più di seimila anni.

Come tutte le innovazioni le religioni hanno costi e benefici. Non è probabilmente una coincidenza che la pubblicazione del libro di Lamarck "Philosophie zoologique" si ebbe dopo che la rivoluzione francese aveva allentato alcuni dei freni che avevano rallentato lo sviluppo della scienza.

Ovviamente la credulità e la predilezione per le leggende sull'origine dell'uomo, diffuse in tutte le popolazioni della terra, non sono confinate a questo solo argomento.

Specialmente quando sono sostenute da interessi politici, queste credenze durano ben al di là dell'età dell'innocenza.

Uno dei problemi della comunicazione attraverso una lingua è la sua ambiguità ed il fatto che una lingua è soggetta a cambiamenti nel tempo. Le parole sono spesso

usate metaforicamente e questo aumenta la confusione. L'ostacolo principale all'accettazione dell'evoluzione viene individuato nell'indicazione all'inizio del Genesi che la terra fu creata in sei giorni, un fatto preso alla lettera da numerose chiese cristiane fondamentaliste.

Vi sono unità di misura basate sul senso comune per tutte le importanti quantità come per esempio il tempo, la lunghezza, la superficie, il volume ed il peso, ma il loro significato varia considerevolmente con il luogo e con l'epoca storica, così come i termini usati per riferirsi ad esse. Per ovvie ragioni il tempo è misurato sulla base di intervalli ciclici, tra cui il giorno è quello più preciso, mentre il mese e l'anno sono più difficili da definire data il loro legame con la localizzazione geografica.

Il "giorno" dell'inizio del Genesi è chiaramente una metafora poetica per un tempo molto più lungo, equivalente a ere geologiche della scienza di durata non specificata. Dopo tutto che cosa è un giorno per un Dio atemporale? In altre parti della Bibbia la parola usata per gli anni dell'età dei patriarchi sembra avere voluto indicare periodi in mesi. L'uso dei termini per misurare il tempo non è l'unico caso in cui l'ambiguità linguistica della Bibbia ha creato grandi conflitti.

LA FONDAMENTALE IMPORTANZA DELLA DEMOGRAFIA PER COMPRENDERE E MISURARE L'EVOLUZIONE.

I miei interessi di ricerca si sono centrati intorno all'evoluzione, e gli organismi su cui posso reclamare una qualche conoscenza sono l'uomo e i batteri. E' ben noto che l'attenzione di Darwin fu attirata sulla demografia dall'enfasi posta da Malthus sulle semplici conseguenze della crescita esponenziale degli organismi viventi. Tutte le entità che si riproducono come gli organismi viventi dopo una fase di crescita esponenziale raggiungono condizioni limite che frenano l'accrescimento.

Uno dei vantaggi di studiare l'uomo è che, nella nostra specie, l'analisi demografica è più facile da fare che in qualsiasi altra specie. Tre dei maggiori fattori dell'evoluzione si possono descrivere quantitativamente in termini demografici. La selezione naturale è una questione di differenza da parte di tipi diversi per caratteristiche ereditabili, nella probabilità di sopravvivenza fino alla riproduzione e nella fertilità (il numero di figli nati). Il teorema fondamentale di Fisher (1930) usa queste quantità per il calcolo di quella che è chiamata la "fitness" Darwiniana di tipi diversi per caratteristiche ereditabili. Casi speciali come la selezione dipendente dalla densità o la fitness inclusiva (che considera anche relazioni di parentela diverse da genitori e figli) richiedono modificazioni del Teorema fondamentale

Anche la deriva genetica (Drift) è pure una questione di individui che si riproducono in una popolazione e nella variazione nel numero dei figli, indipendente dalla selezione naturale. La Migrazione è un'altra quantità di preta

natura demografica, sebbene per i genetisti sia importante distinguere tra due tipi di migrazione (LL Cavalli-Sforza 1973).

Il Tipo 1 è migrazione individuale, vale a dire (almeno storicamente) limitata a distanze brevi e legata allo spostamento, causato dal matrimonio, di almeno uno dei due sposi. Questa migrazione fa diminuire la variazione tra popolazioni causata dal Drift o dalla differenza in selezione naturale in località con caratteristiche ambientali diverse. Il tipo 2 è migrazione di gruppo che porta alla colonizzazione di nuovi territori, che può avere effetti opposti, dal momento che può far aumentare il drift a causa del così detto "effetto del fondatore" (le conseguenze del numero molto ridotto degli effettivi della popolazione iniziale). Le ridotte dimensioni della popolazione, per almeno un certo numero di generazioni, creano le condizioni per l'azione del drift. Inoltre, la selezione naturale, esercitata dalle nuove condizioni ambientali in cui viene a trovarsi la popolazione, può produrre cambiamenti nei patterns della variabilità genetica della popolazione

Il quarto fattore dell'evoluzione, la mutazione, sorgente di tutte le novità evolutive, se considerata su grandi numeri, può pure essere descritta in termini demografici, sebbene sia il fattore per cui è difficile avere buone stime data la sua rarità, che richiederebbe di analizzare un numero di osservazioni molto più grande di quelle che siamo preparati ad usare. L'analisi genomica potrebbe d'ora in avanti, fornire basi precise per la sua misurazione.

La cinetica dei cambiamenti della popolazione per effetto di questi quattro fattori dell'evoluzione è l'obiettivo principale della teoria matematica sviluppata

specialmente da R.A. Fisher, J.B.S. Haldane e Sewall Wright negli anni venti e trenta del 900. Divenne rapidamente uno stimolo di scoperte della matematica nel campo dei processi stocastici, a cui Fisher, Wright e Kimura diedero grandi contributi. La teoria ha continuato a svilupparsi ed è divenuta matematicamente sempre più complessa.

Ci stiamo avvicinando al tempo in cui saranno disponibili dati che permetteranno, usando il lavoro teorico fin qui sviluppato, di stimare l'importanza relativa dei diversi fattori dell'evoluzione in un gran numero di situazioni reali e di sviluppare ulteriormente la teoria generale dell'evoluzione insieme all'epidemiologia genetica.

LA FREQUENZA DEI POLIMORFISMI.

L'analisi della variazione genomica individuale, che ha già rivelato una elevata frequenza dei polimorfismi, è certamente uno dei più importanti prossimi passi dello sviluppo della ricerca. La limitazione principale è l'elevato costo dell'analisi dei genomi individuali. La disponibilità di strumenti di analisi relativamente poco costosi ha indicato che molti siti nucleotidici (nell'ordine dei milioni) sono polimorfici. Fino a che il costo di uno studio genomico individuale completo non si ridurrà ancora, dovremo accontentarci dei polimorfismi che conosciamo oggi, tutti abbastanza frequenti da essere rilevabili in campioni di poche centinaia di individui. Rimarrà tuttavia ancora vero che i polimorfismi più comuni riguardano un solo

nucleotide, e quelli già noti indicano che milioni di siti nucleotidici sono polimorfici, circa uno ogni mille dei circa 3.14 miliardi di nucleotidi che formano il genoma dell'uomo. La validità delle stime disponibili della distribuzione dei polimorfismi è ancora piuttosto approssimata a causa di "vizi di accertamento".

I siti nucleotidici che sono testabili oggi sono stati trovati polimorfici sulla base di campioni abbastanza piccoli di individui su cui sono stati selezionati. I chips disponibili al momento testano un gruppo di polimorfismi scelti in modo in qualche modo arbitrario dal momento che i criteri di selezione non sono sempre noti. Pur tenendo presenti queste limitazioni, è entusiasmante sapere che siamo in grado di testare oltre un milione di polimorfismi nel genoma umano. Naturalmente quando saremo in grado di esaminare un maggior numero di genomi individuali, il quadro generale del polimorfismo nell'uomo potrà divenire molto più complicato ma anche molto più ricco di informazione.

Un notevole risultato è che la gran parte dei polimorfismi sono bi-allelici. Una ulteriore importante caratteristica del DNA mitocondriale e del cromosoma Y è che più del 90% dei siti è costituito da transizioni dovute a mutazioni $A \leftrightarrow G$, o $T \leftrightarrow C$, mentre nel DNA autosomico la frequenza relativa delle trasversioni è maggiore e un po' più simile a quella delle transizioni. Il biallelismo è abbastanza frequente, mentre il tri-allelismo è relativamente raro. Questo è in accordo con la semplice ipotesi che le transizioni abbiano maggior probabilità di accadere dal momento che esse generano un minor livello di cambiamenti chimici, anche se una

spiegazione veramente convincente va ricercata attraverso il chiarimento dei dettagliati meccanismi biochimici coinvolti.

Le differenze nel rapporto tra transizioni e transversioni tra DNA mitocondriale e DNA del cromosoma Y da una parte e quello autosomico dall'altra è probabile rifletta le differenze tra trasmissione uni e bi-parentale. Il DNA nella trasmissione uniparentale ha un tasso evolutivo quattro volte più alto sotto l'influenza del drift, dal momento che la variabile chiave nel determinare il drift, l'effettiva dimensione della popolazione, è quattro volte più piccola che negli autosomi.

Un'altra caratteristica notevole delle mutazioni, che fu per la prima volta messo in evidenza dalla vicenda della "Eva africana", è il fatto che i mitocondri di tutti gli esseri umani viventi discendono da una singola donna, non perché ci sia mai stata una sola donna in esistenza in un dato momento, ma perché i mitocondri di tutte le altre donne che erano vive allora erano differenti e distinguibili da quelli di "Eva" grazie ad una o più mutazioni che avvennero nella madre di "Eva" (o in antenati prossimi nella linea materna che non lasciarono nessun altro discendente oltre ad "Eva" ed ai suoi discendenti) e che non erano presenti negli altri DNA mitocondriali che esistevano al tempo di "Eva" ma sono oggi estinti. Nel famoso lavoro del gruppo di Allan Wilson (Cann et al., 1987) la data di nascita di "Eva mitocondriale" venne stimata tra 150,000 e 300,000 anni fa.

Né la determinazione dell'origine di "Eva" né la sua localizzazione in Africa erano stati rigorosamente provati al momento della pubblicazione, ma ricerche molto

condotte molto dopo hanno fornito solide basi ad entrambe e l'origine africana dell'uomo moderno é stata anche confermato dall'archeologia. La stima corrente del tempo in cui avvennero le mutazioni che distinguono "Eva" é molto vicino all'estremo recente della stima originale 150000 anni con un errore standard del 10%.

Il cromosoma Y, l'altro caso di trasmissione uniparentale, ha prodotto una stima di circa 100000 anni. E' stato suggerito dal gruppo di Barbujani (Dupanloup et al, 2003) che la differenza tra le origini di Adamo e di Eva é conseguenza della poligamia ed, in effetti, il rapporto delle due date non è lontano, come atteso, dal rapporto del numero di mogli per uomo che si osservano ancora oggi in molte parti del mondo.

Che tutti i cloni di questi cromosomi trasmessi uniparentalmente che esistono oggi discendano da un cromosoma antenato che esisteva nel passato, e che tutti gli altri cloni in esistenza al tempo dell'origine dell'unico tipo che sopravvive oggi siano estinti, e' una semplice conseguenza del drift dovuto alla variazione del numero di figli per individuo. Questo non si può facilmente estrapolare agli altri cromosomi, agli autosomi ed al cromosoma X o a parti di essi perché il crossing-over mescola il DNA degli omologhi e distrugge la sequenza originale dei nuclotidi. Brevi segmenti di DNA che non avevano crossing over seguono la stessa regola ma non e' facile distinguere gli estremi di tali segmenti. La maggior parte dei geni regolari sono troppo lunghi per poter essere definiti segmenti brevi. Quando siamo in grado di individuare gli estremi di segmenti che non sono stati oggetto di

crossing over dal momento in cui avvennero le mutazioni note nel singolo nucleotide, ci possiamo aspettare che il tempo di origine di questi brevi segmenti di DNA autosomici sia, in media, quattro volte più grande (e tre volte per il cromosoma X) di cromosomi aploidi trasmessi in modo uniparentale.

L'età media di questi brevi segmenti non ricombinati di DNA in cromosomi trasmessi in modo uniparentale è probabile che sia maggiore di 500,000 anni a causa del fatto che una ignota, ma, come vedremo in seguito, potenzialmente grande proporzione di essi potrebbe essere sottoposta a selezione naturale dovuta al vantaggio dell'eterozigote.

Il vantaggio dell'eterozigote, se non dovuto al segmento stesso può derivare dall'associazione con un segmento vantaggioso vicino ("sovradominanza associativa attraverso il linkage disequilibrium) (Ohta e Kimura, 1970). In effetti, nell'HLA vi sono polimorfismi che possono avere 15 milioni di anni ed esempi di mutazioni con una età evolutiva di alcuni milioni di anni hanno cominciato ad apparire in letteratura. Sotto l'azione del solo drift vi è naturalmente una grande variazione nei tempi individuali di evoluzione (Kimura 1970), ma una età di due milioni contro una attesa di 500000 anni sarebbe una osservazione osservabile solo molto raramente, ed inoltre vi sono altre ragioni per sospettare il vantaggio dell'eterozigote in molti polimorfismi.

Alcuni antropologi sostengono ancora una vecchia ipotesi antropologica, nota come teoria multiregionale secondo cui la variabilità genetica che osserviamo oggi suddivisibile in quattro o cinque razze abbia avuto origine almeno 1.5 milioni di

anni fa, al tempo della prima espansione di Homo erectus, e non sia stata completamente distrutta in seguito.

Questi antropologi sperano di salvare la teoria multiregionale mostrando che alcuni dei geni oggi polimorfici nell'uomo abbiano origini antiche, lontano dall'Africa (una impresa difficile). Ma questo approccio non ha grande probabilità di essere di notevole aiuto. E' naturalmente perfettamente plausibile che mutazioni originatesi in Asia che avvennero prima dell'ultima espansione possano essere state trasmesse in modo da raggiungere l'Africa anche in tempi remoti, e abbiamo esempi di simili diffusioni a ritroso anche se avvenute in tempi posteriori alla più recente espansione originatasi in Africa.

L'UOMO COME SPECIE COSMOPOLITA ED INVASIVA.

Altre complicazioni nell'applicazione della genetica di popolazione allo studio della nostra storia evolutiva vengono dalle contrazioni (colli di bottiglia) e successive espansioni a cui sono state soggette le dimensioni delle popolazioni umane. Tali eventi sono in generale prodotti da grandi innovazioni culturali permesse dal continuo aumento delle nostre capacità di comunicazione e di inventare che hanno enormemente aumentato la rilevanza della evoluzione culturale rispetto ad altre specie anche quelle filogeneticamente vicine.

Questo si riflette nell'aumento, testimoniato dall'evidenza archeologica, del volume del nostro cervello che è adesso circa quattro volte le dimensioni del cervello dell'antenato comune con il nostro cugino più vicino lo scimpanzé. Al contrario le dimensioni del cervello dello Scimpanzé non sono cambiate molto. La separazione avvenne 5-6 milioni di anni fa in qualche posto in Africa, ed un grande numero di specie si originò da quella separazione specialmente nei rami non umani in cui al contrario solo la nostra specie (*H. sapiens sapiens*) sopravvive. La sua notevole omogeneità genetica testimonia una origine recente di cui vi sono evidenze archeologiche che le attribuiscono al massimo un'età di 150000 anni ed un'origine nell'Africa orientale.

Le nuove specie del ramo umano che si sono originate nella prima metà del tempo che ci separa dalla separazione dallo Scimpanzé sono classificati in un genere diverso dal nostro, *Australopithecus* (così chiamato perché trovato principalmente in Africa meridionale). È stato stabilito che il nostro genere, *Homo*, ebbe inizio tra 2.5 e 3 milioni di anni fa, e fu chiamato *habilis*, perché mostrò le prime evidenze di capacità di produzione di utensili anche molto rozzi. Circa 1.7 milioni di anni fa gli utensili erano molto aumentati in numero ed in complessità, permettendo alla specie *H. erectus* di diffondersi dall'Africa all'intera Eurasia. Molto probabilmente il controllo del fuoco (la prima evidenza archeologica è di 1.6 milioni di anni fa) di ovvia utilità per moltissimi usi fu pure di aiuto a sostenere quella espansione. In quei tempi molte specie si formavano per speciazione allopatrica, ma tutte le

specie australopithecine e le specie del genere Homo tranne la nostra sono estinte da molto tempo.

La successiva grande espansione, questa volta all'intero mondo, avvenne molto più tardi e fu iniziata da una popolazione piccola - dati genetici indicano che probabilmente era costituita da solo un centinaio di individui - che viveva in Africa orientale. Le stime del tempo di evoluzione indicano che ci potrebbe essere stato un inizio lento nella crescita e nella espansione forse 100000 anni fa, ma vi fu poi, circa 50-60000 anni fa, una notevole accelerazione dell'espansione, sostenuta da evidenza archeologica, verso l'Asia e oltre fino ad interessare l'intera terra. A quell'epoca il linguaggio era completamente sviluppato e deve essere stato uno degli strumenti principali della sofisticata comunicazione che consentì l'accumularsi e il diffondersi delle acquisizioni culturali. Il linguaggio è certamente la differenza principale tra uomini ed animali.

Le tribù sono (o erano in origine) essenzialmente gruppi sociali che condividevano una lingua e una identità culturale. La lingua parlata dalla tribù ancestrale dell'Africa orientale, quando iniziò a crescere e ad espandersi, è molto probabilmente quella da cui derivano le circa 6000 lingue parlate oggi nel mondo. I linguisti trovano difficoltà nell'accettare che vi sia stata una sola lingua in origine perché l'evoluzione linguistica è molto rapida ed ha prodotto profonde differenze tra le lingue parlate oggi. L'evidenza principale in favore di una lingua originaria unitaria è che, con l'eccezione di qualche raro individuo linguisticamente disabile, tutti possono imparare una qualsiasi lingua oggi esistente.

Altre innovazioni importanti che favorirono l'espansione furono "invenzioni" quali la produzione di un sempre più complesso ed utile insieme di strumenti e di procedure, quali la navigazione, che permise di superare tratti di mare come quelli che separano dalle altre terre la Nuova Guinea e l'Oceania che furono raggiunte oltre 40000 anni fa.

Circa 13000 anni fa (o poco dopo) ebbe termine l'ultima di una serie di grandi glaciazioni e i ghiacci iniziarono a ritirarsi. Il cambiamento climatico associato con questo evento favorì la diffusione della vegetazione erbacea in aree quali il medio oriente in cui poi si sviluppò l'agricoltura. Il cambiamento della vegetazione causò cambiamenti nella fauna e anche sull'uomo. La popolazione umana totale della terra al tempo dell'origine dell'agricoltura è stimata tra 1 e 15 milioni di individui, quindi circa un migliaio di volte minore di quella attuale ed un migliaio di volte superiore di quella presente prima dell'inizio dell'ultima espansione. L'espansione fu aiutata dall'innovazione e, spingendo l'uomo in aree geografiche nuove, fu di stimolo per lo sviluppo tecnologico che portò, per esempio nuove armi come l'arco e la freccia, l'uso del bambù (Asia orientale) e l'uso di pelli e fibre per affrontare i climi freddi. Specialmente alla fine della glaciazione vi fu grande innovazione negli strumenti di pietra che divennero più piccoli e più diversificati (in Europa questa industria è nota come mesolitico).

Ma l'economia era ancora basata sulla caccia e la raccolta anche se, grazie all'innovazione tecnologica, la popolazione nelle aree climaticamente più favorevoli stava crescendo. In queste aree l'archeologia ci fornisce evidenza del sorgere,

indipendentemente in diverse parti del mondo, di quella che sarà l'importantissima innovazione dei millenni successivi: la produzione di cibo, cioè le coltivazioni e l'allevamento degli animali. L'agricoltura, basata sulla conoscenza dei meccanismi della vita (la riproduzione e la possibilità di modificare le caratteristiche delle piante e degli animali) sostituiva la raccolta del cibo dall'ambiente naturale.

La produzione del cibo ebbe inizio indipendentemente nelle varie aree, nel senso che le piante e gli animali che furono addomesticati erano locali, già precedentemente usati come cibo dai cacciatori e raccoglitori (i pesci sono stati allevati in modo equivalente solo recentemente). Tra le aree principali in cui ebbe origine l'agricoltura la più antica si trova tra il Medio Oriente e la Turchia. Il sito di Abu Hureyra vedeva praticati, già 11500 anni fa, l'allevamento e l'agricoltura. La domesticazione di piante ed animali locali ebbe inizio un po' dopo in Cina (indipendentemente nel Sud e nel Nord), nel Sahara che non era un deserto all'epoca, in Nuova Guinea, nell'altopiano messicano e nell'Africa occidentale.

Dai centri di origine l'agricoltura si diffuse lentamente alla velocità di 1-2 Km all'anno, portata sia dalla migrazione degli agricoltori che cercavano nuove terre alla periferia dell'area in cui era praticata, sia dall'imitazione da parte dei cacciatori e raccoglitori che imparavano la sua pratica dai loro vicini, con cui probabilmente c'erano anche scambi matrimoniali. L'agricoltura produsse notevoli cambiamenti nella dieta dell'uomo. Le prime piante coltivate prodotte nelle diverse parti del mondo (grano, mais e riso) sono ancora oggi le più importanti del mondo. L'agricoltura fece diminuire il nomadismo associato con le attività di caccia e

raccolta anche se, in tempi successivi, in zone semi aride lo sviluppo di pratiche di pastorizia quasi esclusiva di altre forme di produzione di cibo ha generato nuove forme di nomadismo dalle caratteristiche particolari.

La sedentarietà e l'aumento della densità della popolazione causò una generalizzata stratificazione della società, la differenziazione delle attività lavorative, l'inizio di attività di governo, l'opportunità di possedere la terra (anche in grandi estensioni) e l'accumularsi del potere nelle mani di pochi, tutte cose impossibili in una società basata sulla caccia e la raccolta. In tempi ancora successivi il proseguimento di questo processo pose le basi per ulteriori innovazioni come quella dell'uso dei metalli. Il primo fu il bronzo (circa 5000 anni fa), mentre circa 1500 anni dopo si sviluppò l'uso del ferro, in un'area simile, al confine tra Europa ed Asia, a nord del Caucaso. Nella stessa area avvenne la domesticazione del cavallo. Insieme, i metalli ed il cavallo crearono le condizioni per l'instaurarsi di uno stato di guerra tra vicini che vide quelli che prima erano cacciatori trasformarsi in guerrieri. Il gran numero di prigionieri prodotti dalle guerre alimentò una vasta classe sociale costituita da schiavi. Divenne possibile per tribù numerose occupare militarmente aree ritenute più favorevoli. I guerrieri a cavallo originari delle steppe della Mongolia iniziarono la conquista dell'Asia. Più tardi, la domesticazione del cammello consentì agli Arabi di conquistare la Turchia ed il Nord Africa, una espansione che, molto dopo (nel settimo secolo dopo Cristo), si estese alla Sicilia e alla Spagna, mentre i mercanti arabi si spingevano nell'Africa orientale e meridionale. Alcune invenzioni ed innovazioni si sono rivelate

estremamente importanti e la loro influenza sulla nostra vita, inclusa la nostra genetica, continua a crescere ad un ritmo sempre più rapido. La storia culturale ha quindi iniziato a dominare la struttura genetica della nostra specie e ad influenzare il destino di molte altre specie, come conseguenza della nostra alterazione dell'ambiente (il confronto tra osservazioni archeologiche e ed evoluzione genetica dell'uomo moderno si possono trovare in "L.L.Cavalli-Sforza et al; 1988, 1994, L.L.Cavalli-Sforza e M.W.Feldman (2003)").

IL CONTRIBUTO DELLA MUTAZIONE E DELLA SELEZIONE AL POLIMORFISMO.

Dei quattro fattori dell'evoluzione, la mutazione è responsabile dell'origine dei polimorfismi, la selezione per il cambiamento della loro frequenza in risposta alle condizioni ambientali. Nei diploidi la selezione sotto forma di eterosi favorisce la stabilità dei polimorfismi. Al contrario il drift, sia nel breve che nel lungo periodo, tende a eliminare i polimorfismi e a differenziare le popolazioni tra di loro, mentre la migrazione individuale tende a ridistribuire i polimorfismi esistenti all'interno della specie. Kimura (1968) ha dimostrato che nell'evoluzione molecolare la maggior parte delle mutazioni sono neutrali in termini di selezione, un quadro in cui la selezione ed il drift sono le cause principali delle differenze tra specie. Cosa possiamo dire dei milioni di polimorfismi che non mostrano nessuna gran differenza nella nostra specie almeno tra le quattro popolazioni fin qui testate dal programma del Genome Institute for HapMap del National Institute for Health? Vi

sono naturalmente alcuni geni, o più precisamente, alcuni segmenti che mostrano forti differenze tra popolazioni, un segnale di selezione differenziale che si estende ai segmenti vicini a causa del linkage disequilibrium, ma la maggior parte dei polimorfismi mostrano poca differenza tra le quattro popolazioni, confermando la grande omogeneità genetica della nostra specie se confrontata con la maggior parte dei Mammiferi.

Un domanda interessante e' perché vi siano così tanti polimorfismi. Viene in mente la seguente situazione:

- 1) I polimorfismi erano presenti nella specie prima della sua origine, dovuti a selezione differenziale che portò gli alleli alla frequenza che osserviamo oggi, ma che ha da allora cessato di agire, così che oggi gli alleli sono neutrali e fluttuano solo sotto l'influenza del drift, in equilibrio con la migrazione.
- 2) I polimorfismi sono stati portati alla loro frequenza media attuale dall'associazione con geni vicini sul genoma che sono (o sono stati) sotto selezione e le differenze tra popolazioni che vediamo sono dovute al drift in equilibrio con la migrazione.
- 3) Le frequenze dei polimorfismi possono essere il risultato di due diverse frequenze di mutazione in assenza di altre pressioni, anche se i cambiamenti in frequenze geniche solo sotto pressione di mutazione sono estremamente lenti. Inoltre il bi-allelismo di quasi tutti i siti nucleotidici ci indica che siamo lontani da un equilibrio con i tassi di mutazione. Dovremmo aspettarci in questo caso a

molti siti con più di due nucleotidi presenti alternativamente. In effetti, il tempo richiesto perché si instauri un equilibrio sotto un tasso di mutazione u in una popolazione grande è dell'ordine di $1/u$ generazioni. Abbiamo scarse conoscenze sui tassi di mutazione, e per pochissimi geni; le stime per i valori medi dei tassi di mutazione sono per lo più derivati dall'analisi evoluzionistica, ma sappiamo anche che alcuni singoli tassi di mutazione possono essere molto diversi dal valore medio. In generale l'ordine di grandezza dei cambiamenti delle frequenze geniche dovuti solo a pressione di mutazione è dei milioni di generazioni. Vi sono forti ostacoli all'osservazione di equilibri determinati solo dalla mutazione, come gli eventi accidentali che possono indurre un forte effetto del drift, come nel caso di forti effetti del fondatore, quali quelli che generano nuove specie, oppure una forte selezione in ambienti fortemente differenziati.

- 4) Il vantaggio selettivo degli eterozigoti genererà equilibri polimorfici stabili. Anche se un allele è letale o quasi frequenze di eterozigosi vicine al 20 o 30% sono state osservate all'equilibrio, per esempio per l'anemia falciforme. Il vantaggio dell'eterozigote si estende a siti neutrali molto vicini sul genoma e quindi strettamente associati. La velocità a cui si raggiunge tale equilibrio è molto maggiore che sotto pressione di mutazione. Nel primo esempio di vantaggio dell'eterozigote, l'anemia falciforme, il coefficiente di selezione dell'ordine del 10%. Questo vuol dire che la fitness dell'omozigote normale era del 10% più piccola di quella dell'eterozigote, ed era sufficiente a mantenere una frequenza del 20% di eterozigoti all'equilibrio, anche se l'altro omozigote era

essenzialmente letale. Vi sono pochi altri esempi di vantaggio dell'eterozigote altrettanto chiari dell'anemia falciforme (o della talassemia che è molto simile). Con una selezione di questa forza sono necessarie circa 100 generazioni per portare una popolazione di 10000 individui dalla comparsa della mutazione vicino ad una condizione di equilibrio. Il tempo in generazioni per rimpiazzare di un nuovo mutante è proporzionale al reciproco del coefficiente di selezione. Così, nel corso dell'ultima espansione dell'uomo (2000 x 30 generazioni, circa 60,000 anni) una nuova mutazione che fosse apparsa all'inizio e avesse mostrato vantaggio dell'eterozigote potrebbe essere vicina all'equilibrio oggi, anche con un coefficiente di selezione di solo lo 0.5%. Se l'equilibrio dovuto al vantaggio dell'eterozigote esisteva anche prima dell'inizio dell'espansione, il coefficiente di selezione potrebbe essere molto più piccolo, anche solo 1/10000. Un coefficiente così piccolo richiede naturalmente un tempo proporzionalmente più lungo per far raggiungere l'equilibrio al polimorfismo (una introduzione numerica semplice a questo tipo di calcolo si può trovare in Cavalli-Sforza e Bodmer 1999).

Il modo più sicuro per testare l'esistenza del vantaggio dell'eterozigote è dimostrare che vi è deviazione dall'equilibrio di Hardy Weinberg negli adulti, con eccesso di eterozigoti, anche se fu necessario un campione di mille individui per dimostrarlo nel caso dell'anemia falciforme. Per coefficienti di selezione più piccoli le dimensioni del campione diventano proibitive. Tuttavia un'analisi di 81 geni, studiati a livello del genoma, in cui la selezione mostrava effetti importanti,

identificati dalle notevoli differenze tra alcune popolazioni nelle frequenze geniche di polimorfismi che si trovano in (o vicino a) segmenti di DNA con funzioni speciali, che ci sono (almeno in parte) note, mostrò che circa un quarto di essi era in grado di attribuire un vantaggio all'eterozigote (eterosi). E' probabile che l'eterosi sia coinvolta in un numero anche più grande di polimorfismi, tenendo conto del fatto che la stabilizzazione di un polimorfismo si estende anche ai geni neutrali vicini attraverso il linkage disequilibrium. Naturalmente, questo effetto stabilizzante, dovuto al linkage disequilibrium con varianti sotto selezione, è tanto più piccolo tanto più debole è la selezione (Ohta e Kimura 1970).

COLLI DI BOTTIGLIA DEMOGRAFICI SEGUITI DA CRESCITA DEMOGRAFICA.

I Colli di bottiglia demografici sono particolarmente importanti nel determinare la struttura genetica di una popolazione, specialmente se sono seguiti da una notevole crescita demografica.

Sappiamo che la nostra specie ha avuto una storia di questo tipo al tempo della sua origine, negli ultimi 100.000 anni, con espansioni limitate quasi solo all'Africa per i primi 50.000 anni, seguite da un'espansione al mondo intero. Più tardi vi furono alcune espansioni di notevoli dimensioni a cominciare dal tempo della scoperta dell'agricoltura circa 10000 anni fa. Più locali queste furono responsabili per i quattro o cinque clusters di similarità genetica che corrispondono più o meno

ai continenti. Ma poi ve ne furono numerose altre, geograficamente più limitate. La navigazione transoceanica nell'epoca dopo Colombo portò alla diffusione degli Europei ma anche della diffusione degli Indiani e, in minor misura, dei Cinesi, a parti del mondo che offrirono opportunità economiche interessanti. I costumi sociali hanno avuto la tendenza di mantenere le società abbastanza endogamiche, persino quando gli individui migravano lontano dalla madre patria, sebbene vi siano esempi di tendenza alla fusione come negli Stati Uniti ed in Brasile. Ma in Europa ed altrove i gruppi di colonizzatori mantennero la loro identità culturale originaria, inclusa la lingua, per periodi di straordinaria lunghezza. La lingua Basca è forse discendente, anche se soggetta a notevoli cambiamenti, da una famiglia di linguaggi che erano diffusi dall'Eurasia al Nord America prima di 20000 anni fa, forse ancora prima al momento dell'arrivo dell'espansione proveniente dall'Africa. Ma la maggior parte delle lingue si sono diffuse più recentemente, e di solito hanno quasi completamente rimpiazzato quelle preesistenti, probabilmente durante l'ultimo post glaciale insieme alla diffusione dell'agricoltura.

In tempi più vicini a noi vi è stata una serie di espansioni di minor rilevanza, iniziate con l'età dei metalli. La diaspora degli Ebrei li portò in diversi continenti 2500 e 1900 anni fa e, di nuovo, in tempi più recenti, e le loro comunità crebbero quasi ovunque. Meno di 2000 fondatori, per la maggior parte olandesi, ma con qualche presenza francese e tedesca, si stabilirono a Città del Capo in Sud Africa e si sono accresciuti quasi mille volte in 350 anni e formano adesso circa metà della popolazione bianca del Sud Africa. I Franco canadesi del Quebec hanno avuto

origine quasi nello stesso tempo da circa un migliaio di donne francesi che erano state invitate dal re Luigi XIV a divenire "Figlie del Re", e a ricevere una dote se avessero accettato di sposare i cacciatori di pelli francesi che erano attivi in Quebec. Anche qui si ebbe una crescita di circa mille volte in 350 anni. Queste popolazioni sono state studiate dal punto di vista genetico ed è stato trovato che le rarità genetiche o le malattie presenti in uno o pochi fondatori, a cui si risale attraverso la ricostruzione delle genealogie, vi sono ora relativamente frequenti.

Una parola di cautela: ci si può domandare se i valori numerici relativi ai tempi dei processi evolutivi siano corretti dal momento che sono, di solito, calcolati attraverso la simulazione dei processi di crescita delle popolazioni, per le quali è assunta una crescita esponenziale. E' assai probabile che la crescita della nostra specie, durante la sua espansione dall'Africa, sia stata assai meno che esponenziale e si sia accresciuta linearmente con il tempo. Il miglior modello per una espansione demografica e geografica ("demica" come si dice in breve) è probabilmente una estensione del modello introdotto da Fisher (1937) per una mutazione vantaggiosa, chiamato "l'onda di avanzamento" ("the wave of advance"), che mostra che l'espansione con densità di popolazione costante da luogo ad un tasso di crescita pressoché costante, cioè lineare e non esponenziale rispetto al tempo.

L'EFFETTO DEL FONDATORE IN SERIE NELL'ESPANSIONE DELL'UOMO MODERNO.

L'analisi da parte di Rosenberg et al. (2002) del HGDP con circa 400 marcatori microsatellite ha generato una serie di ulteriori lavori. Prugnolle et al. (2005) hanno mostrato che l'eterozigosi media delle 52 popolazioni decresceva in modo quasi perfettamente lineare del 16% (da 82% a 64%), in funzione della distanza dall'origine. Noi abbiamo confermato i loro risultati con il doppio di microsatelliti, usando altri 400 marcatori che erano nel frattempo stati pubblicati, ed abbiamo offerto (Ramachandran et al. 2005) una interpretazione che spiegava bene la caduta osservata della variazione genetica misurata come eterozigosi. La simulazione ha usato un modello di "effetto del fondatore in serie" che può essere così riassunto: la popolazione originale dell'Africa orientale si accrebbe da dimensioni No a dimensioni saturanti e quindi generò una colonia ad una certa distanza che a sua volta crebbe a dimensioni di saturazione.

La prima colonia generò quindi una seconda colonia che si spostò ancora nello spazio, per crescere e ripetere il processo fino a che gli estremi della terra erano stati raggiunti. L'origine di ogni colonia era un episodio di effetto del fondatore che si ripete in serie ad ogni nuovo episodio di colonizzazione. La simulazione ha mostrato una caduta lineare della eterozigosità media ad un tasso che corrispondeva bene con quello osservato nei dati dei microsatellite, usando numeri per le tribù e per il modello di crescita che erano antropologicamente ragionevoli.

Secondo questo modello cento eventi di colonizzazione coprirebbero in 40.000 anni i 25.000 km che separano l'origine, arbitrariamente fissata a Addis Abeba dalla colonia terminale dalla punta meridionale dell'America del Sud. Gli eventi di colonizzazione si sono svolti via terra evitando di attraversare gli oceani. Un simile esercizio ad opera di Liu et al (2006) ha suggerito in 300 il numero degli eventi di colonizzazione.

Questi modelli possono essere migliorati ma essi almeno mostrano che una simulazione che usa il modello di Fisher dell'onda di avanzamento (1935) si adatta pure bene alle osservazioni ed è probabilmente preferibile in quanto corrisponde più da vicino alla realtà antropologica, ma richiederebbe per un adattamento dettagliato, una informazione antropologica più specifica di quella disponibile per le società dei cacciatori e raccoglitori.

LA CORRELAZIONE TRA DISTANZE GEOGRAFICHE E DISTANZE GENETICHE.

L'analisi di Ramachandran et al. (2005) dei dati di Rosenberg et al (2002) ha anche stimato la correlazione tra distanze genetiche e distanze geografiche calcolate su tutti le possibili coppie di popolazioni HGDP. E' stata trovata una correlazione lineare molto alta tra le due distanze ($r=0.89$), una delle più elevate riportate nella letteratura biologica. Si deve sottolineare che le 52 popolazioni HGDP sono tutte indigene, cioè sono localizzate oggi nello stesso posto in cui si

trovavano prima delle migrazioni successivamente a Colombo, così che in America si considerano solo popolazioni native. E' quindi ovvio che le popolazioni che appartengono ad un continente sono più simili tra loro che quelle che appartengono a continenti diversi, ma questo non è sufficiente perché popolazioni di continenti diversi ma geograficamente vicine sono pure più simili geneticamente di quanto ci si aspetterebbe sotto la semplice ipotesi che esse siano parte dei rispettivi continenti di appartenenza, senza considerare le distanze geografiche.

Così per esempio Arabi ed Etiopi, Spagnoli e Marocchini ci si può aspettare siano più simili che due popolazioni Africane, o due popolazioni Europee o due popolazioni Asiatiche prese a caso. Le coppie che si discostano dall'attesa coinvolgono alcune popolazioni Africane particolarmente antiche in termini della loro storia evolutiva, ed hanno avuto, di conseguenza, un tempo più lungo per divergere. Pure lievemente discordanti con l'attesa sono popolazioni che, come atteso, sono note per essere state oggetto di mescolamento anche se parziale. L'interpretazione più semplice di questi risultati, tenendo anche presente dell'effetto di eventi del fondatore in serie, è che la differenziazione genetica osservata per i microsatelliti è in gran parte dovuta all'equilibrio tra drift e migrazione.

Non abbiamo rilevato effetti importanti della selezione naturale, ma osservazioni preliminari con altri tipi di mutanti, o con altri approcci mostrano anche evidenza di selezione naturale direzionale come indicato dalle importanti differenze tra le popolazioni HapMap. Questo riassunto può coprire solo gli aspetti principali di questi risultati. Nel lavoro originale si indica che il drift può dar ragione di almeno il

78% della varianza osservata per i microsatelliti. Viene tuttavia lasciata incertezza sul ruolo, se ve ne è uno, dell'eterosi nel processo.

VARIAZIONE GENETICA TRA ED ENTRO POPOLAZIONI ED IL PROBLEMA DELLA RAZZA.

Lewontin (1981) ha mostrato all'inizio degli anni 80 del secolo scorso che, quando si stima la variazione genetica per marcatori proteici confrontando due individui presi a caso dalla stessa popolazione o presi a caso dal mondo intero, la prima stima risulta assai più grande della seconda. Mentre la variazione entro popolazione è l'85% del totale quella tra popolazioni è solo il 15%. Indagini successive, condotte su un numero limitato di popolazioni ma utilizzando anche altri marcatori, hanno confermato questi risultati. In effetti, i dati di Rosenberg et al. (2002) abbassano la stima di Lewontin fino al 10% o persino al 5%. La stima è quindi perfino inferiore della già piccola stima iniziale. Sappiamo che la stima dipende dalla scelta di popolazioni e marcatori. Ma anche una così ridotta variazione tra popolazioni è certamente sufficiente per la ricostruzione della storia evolutiva delle popolazioni, ma è sufficiente per distinguere le razze in un qualche modo utile? Il confronto con altri Mammiferi mostra che la nostra specie è quasi in fondo alla scala della variazione tra popolazioni. Ma anche così ci sono affidabili metodi statistici che ci consentono di assegnare gli individui alle popolazioni di

provenienza se usiamo un sufficiente numero di marcatori. Ma questo è sufficiente per distinguere le razze? Darwin aveva già una risposta.

Diede due ragioni per dubitare dell'utilità del concetto di razza nell'uomo:

- 1) la maggior parte dei caratteri mostrano una chiara continuità geografica;
- 2) i tassonomi hanno generato una gran varietà di classificazioni delle razze umane. Darwin ne riporta che il numero di razze stimato dai suoi contemporanei varia da due a sessantatré.

Rosenberg et al. hanno analizzato la potenza statistica della suddivisione più efficiente dei loro dati in razze, con un numero di gruppi da due a sei, e mostrano che cinque gruppi hanno una potenza statistica ragionevole. E' da sottolineare che questo risultato è certamente influenzato dalle popolazioni scelte per l'analisi. I cinque gruppi, non molto diversi da quelli proposti da partizioni presenti in letteratura da qualche tempo, sono:

- 1) le popolazioni dell'Africa sub-sahariana,
- 2) le popolazioni Nord Africane, europee e dell'Asia occidentale limitata ad Est dal deserto e dalle montagne dell'Asia centrale,
- 3) le popolazioni del resto dell'Asia orientale,
- 4) le popolazioni dell'Oceania,
- 5) le popolazioni delle Americhe.

Ma quanto buona è questa partizione? L'analisi di Ramachandran et al. degli stessi dati fornisce una predizione delle differenze genetiche tra le stesse popolazioni attraverso il più semplice degli strumenti geografici: le distanze geografiche tra le

due popolazioni. L'analisi di Rosenberg et al aggiunge l'importante conclusione che la classica classificazione in continenti deve essere modificata per rimpiazzare i bordi continentali con barriere geografiche reali: i grandi oceani, o i deserti come il Sahara, o altri deserti e le grandi montagne come quelle dell'Asia centrale. Queste barriere hanno fatto diminuire, ma certamente non completamente soppresso, gli scambi genetici attraverso di loro. Così l'analisi di Rosenberg et al. conferma un pattern di variazione basato su pseudo continenti che non sopprime la continuità di base della variazione genetica, ed, in effetti, un'estensione dell'analisi originale di Rosenberg et al. ha mostrato che le popolazioni che sono geograficamente vicine hanno una notevolissima similarità genetica, ben maggiore di quella suggerita dall'appartenenza ad una ad una partizione pseudo continentale.

La variazione genetica è così assai vicina ad essere continua, come osservato da Darwin. Ma dietro questa continuità vi è un'altra discontinuità che non è ancora stata analizzata. Essa è dovuta alle barriere tra piccoli gruppi sociali, basate su separazioni religiose e stratificazioni socio economiche. Queste creano anche differenze nell'ambiente degli individui e dei piccoli gruppi, attraverso limitazioni nella dieta e nella disponibilità di risorse in generale ed altre auto imposte. Questa analisi suggerisce che è importante riconoscere e studiare gruppi molto più piccoli delle principali "razze" tradizionali. Per chiarire quali gruppi vogliamo arrivare a distinguere, dobbiamo prima rispondere con chiarezza alla domanda: quale scopo abbiamo in mente quando cerchiamo di distinguere le razze?

Il numero dei gruppi da distinguere dipende dalle differenze tra popolazioni che sono realmente utili per qualche valido scopo, e dobbiamo ancora decidere con quali criteri scegliere le popolazioni che vogliamo distinguere. La vera questione quindi rimane: per che motivo abbiamo bisogno delle razze? Iniziamo dal metterci d'accordo su quella che potrebbe essere la principale ragione per definire le "razze".

Per inciso, personalmente preferirei abbandonare del tutto la parola "razza" per la sua connessione con odiosi episodi di razzismo con cui ci dobbiamo continuamente confrontare. La parola popolazione è utile in statistica per definire il gruppo da cui noi otteniamo un campione e forse il termine più neutrale potrebbe essere quello usato da Rosenberg et al "gruppi" (clusters)". Vogliamo definire gruppi genetici utili.

Ma più che un termine da usare quello che cerco è un accordo generale su una buona ragione per fare ricerca che mira a definire una stratificazione utile delle popolazioni, e mi pare che possiamo trovarlo in una ricerca utilizzabile dalla medicina, per la diagnosi e la terapia.

DOVREMO FARE DIAGNOSI E TERAPIA SU MISURA PER CIASCUN GENOMA

INDIVIDUALE?

Oggi si potrebbe obiettare che la risposta finale è comunque un'altra: non abbiamo bisogno di nessun tipo di razza neppure per ragioni mediche. Il genoma individuale si appresta ad essere la miglior sorgente di informazione per la diagnosi e la terapia, anche considerando la forte variazione individuale nella risposta a molti farmaci.

Ma quanto dovremo aspettare prima che questo si avveri? Inoltre le malattie più importanti (le malattie complesse) non si comportano come semplici tratti mendeliani. Sono complessi in due modi. Uno è che probabilmente sono poligeniche, cioè vi è probabilmente una grande varietà di geni che contribuiscono a determinare la patologia. Specifiche malattie mendeliane sono spesso causate da differenti geni e spesso polimorfismi diversi entro lo stesso gene causano differenti patologie, e in malattie poligeniche, quasi ogni caso individuale o pedigree può avere un suo specifico sistema poligenico. L'altra è che l'ambiente causa variazione individuale addizionale e gli effetti ambientali individuali sono di solito scarsamente conosciuti.

E' chiaro che la conoscenza del genoma sarà molto utile per estendere le nostre conoscenze sulle malattie mendeliane e sulle risposte individuali ai farmaci. Ma quando si viene alle malattie più difficili, le malattie complesse, l'ambiente è pure molto importante, ed è chiaro che dobbiamo essere pronti a tenere conto

dell'importanza della stratificazione entro il gruppo, la maggior parte della quale può essere dovuta a differenze socio economiche piuttosto che genetico.

Noi naturalmente speriamo che il genoma sarà pure utile per comprendere meglio il background genetico delle malattie complesse (sempre che i contributi ambientali siano accuratamente studiati allo stesso tempo). Su questo in ogni caso forse dovremo aspettare per sviluppi completamente nuovi sul genoma formato dal DNA delle cellule germinali, che è quello che conta per l'ereditarietà. Sarà anche necessario studiare cosa succede in cellule somatiche, cioè l'epigenetica. Così, perché la conoscenza del genoma dia risposte a tutte o alle più specifiche domande rimane molto lavoro da fare, e forse la sola ragionevole risposta alla domanda "quanto ci vorrà?" è che speriamo ci siano importanti progressi in quest'area in questo secolo.

Anche così, per molte di queste indagini può ancora essere necessario studiare gruppi di individui per risolvere specifici problemi, ma dovranno essere popolazioni più piccole che una per continente. L'epidemiologia genetica mostra che molte malattie mendeliane sono concentrate in alcuni gruppi sociali di solito piccoli specialmente per le malattie più rare.

La ragione è semplice per le malattie recessive: le recenti espansioni demografiche che sono avvenute negli ultimi secoli o negli ultimi millenni in tante parti del mondo hanno creato gruppi di casi di specifiche malattie genetiche, originate da una o poche mutazioni che si sono verificate nel corso della crescita

demografica di specifiche popolazioni. Questi gruppi di casi possono essere ancora oggi gruppi sociali coerenti. E sono questo tipo di gruppi che probabilmente sono molto utili per lo studio delle malattie complesse. Il numero di casi di una malattia ben caratterizzata ci può dare una indicazione approssimativa del tempo in cui si sono originate le mutazioni rilevanti per la malattia.

Ma alcune malattie possono essere causate da molte mutazioni - come è chiaro per esempio per la fibrosi cistica. Le dimensioni e l'importanza del gene contribuiscono a rendere la malattia complessa. La storia demografica di una popolazione può aiutare a predire quali gruppi mostreranno una maggiore similarità nelle forme in cui si presenta la malattia.

La notevole differenza nelle patologie genetiche trovate tra gli Ebrei di origine nord europea e mediterranea (Aschenaziti e Sefarditi) indica che i cluster genetici di individui possano dover essere assai piccoli per essere utili: anche un gruppo relativamente piccolo come gli Ebrei deve essere ulteriormente suddiviso per la ricerca in genetica medica. Se le dimensioni di questi gruppi fossero da prendere come esempi delle dimensioni dei gruppi utili a questi scopi ne deriverebbe la necessità di definirne oltre un migliaio per l'intera l'umanità.

E' probabilmente legittimo dichiarare che, dal momento che la predisposizione genetica a molte malattie è importante, persino per la risposta ad alcune malattie infettive, riconoscere e studiare gruppi genetici di dimensioni utili allo scopo ci potrebbe aiutare per migliorare le nostre capacità di diagnosi e terapia. In pratica ci

potremmo dirigere verso lo studio di gruppi sempre più piccoli di individui, anche se l'utilizzazione dei genomi individuali potrà richiedere ancora anni di studio per poter essere utilizzata in modo veramente adeguato. Fino ad allora mi sembra che sia utile, insieme all'esplorazione di nuovi differenti approcci, ampliare e migliorare le raccolte del tipo del HGDP.

IL PROGETTO HUMAN GENETIC DIVERSITY PROJECT (HGDP) E IL SUO FUTURO.

Nature Reviews-Genetics ha pubblicato nell'Aprile del 2005 una breve storia del HGDP-CEPH (LL Cavalli-Sforza, 2005). La collezione esistente è costituita da 1064 culture di lymphociti B da 52 popolazioni. Tutti i continenti, con l'eccezione dell'Australia, sono rappresentati ma inevitabilmente non nel modo che era stato originalmente auspicato, con popolazioni distribuite regolarmente nello spazio. Vi è un certo raggruppamento spaziale delle popolazioni dal momento che alcuni paesi (Cina, Israele, Pakistan) sono stati particolarmente attivi e generosi nel fornire campioni. Il progetto è stato reso possibile dalla gentilezza dei ricercatori che hanno donato le linee cellulari che avevano raccolto, e dal fatto che CEPH ha reso disponibile l'attrezzatura necessaria per far crescere le linee cellulari, produrre il DNA e prepararlo per la distribuzione ai laboratori. L'attrezzatura era stata messa insieme precedentemente per produrre le mappe cromosomiche umane usate per

studi di associazione sulle genealogie dello Utah e su altre. In questo modo HGDP-CEPH ha richiesto fondi molto limitati.

L'idea originale del progetto HGDP era di raccogliere 10,000 linee cellulari, 25 individui da ciascuna di 400 popolazioni. Il progetto Genographic progetta di raccogliere materiale da 100 individui di ciascuna di 1000 popolazioni, ma il materiale raccolto sarà solo DNA dalla saliva e, in una piccola proporzione di casi, campioni di sangue. Sfortunatamente non produrrà linee cellulari. Con la tecnologia esistente l'analisi di 1000 individui è ancora un po' impegnativa per un singolo laboratorio a meno che sia limitata a marcatori particolari, come nel caso dei microsatelliti, che in ogni caso hanno richiesto notevole impegno. Vi è un inevitabile compromesso tra numero di marcatori e numero di individui che analizzano.

Le culture di B limfociti sono ancora il materiale migliore da raccogliere e conservare. Il DNA rimarrà per un po' di tempo il più importante prodotto della raccolta di genomi, ma le cellule offrono importanti prospettive, per esempio per studi di epigenetica. Le loro proteine e l'RNA possono fornire informazione sui processi epigenetici e sull'evoluzione epigenetica che avviene durante la storia dell'individuo, almeno per quei geni che sono impegnati nella manutenzione e nella riproduzione della cellula e per specifiche funzioni immunologiche che avvengono nei Linfociti B. Di conseguenza queste cellule contengono informazioni sugli aspetti immunologicamente significativi dell'ambiente delle popolazioni.

E' importante mantenere ed estendere la collezione HGDP attuale. Fino ad ora HGDP e' stato sviluppato con investimenti minimi. Vi e' la speranza che si espanda anche se il finanziamento governativo per la ricerca scientifica e' calato quasi ovunque. Chiaramente i paesi in cui gli studi genomici sono destinati a produrre frutti nel breve periodo sono quelli per cui e' ragionevole investire di più in questo momento. E' praticamente inevitabile che i due gruppi più grandi di popolazioni, uno costituito dall'Europa, Nord Africa ed Asia occidentale incluso Pakistan ed India ed uno dall'Asia orientale, saranno studiati in modo preferenziale dal momento che queste due aree, economicamente sviluppate, si presentano come i due poli di una distribuzione continua di popolazioni. Ma sarà importante dedicare una parte dello sforzo alle popolazioni indigene del resto del mondo.

Alternativamente i nuovi immigranti che si spostano nelle parti economicamente favorite del mondo non avranno accesso all'informazione biomedica disponibile agli altri residenti. Non solo l'ineguaglianza è inaccettabile ma in questo modo la nostra comprensione dell'evoluzione dell'uomo rimarrà incompleta.

CITAZIONI BIBLIOGRAFICHE.

1. Cann RL, Stoneking M, Wilson AC 1987. Mitochondrial DNA and human evolution. *Nature* 325:31-6.
2. Cavalli-Sforza LL 1973. Some current problems in human population genetics. *Am.J.Hum.Genet.* 25:82-104.
3. Cavalli-Sforza LL. 2005. The Human Genome Diversity Project: Past, Present and Future. *Nature Reviews/Genetics.* 6 (4): 333-40.
4. Cavalli-Sforza LL, Bodmer WF 1999 *The Genetics of Human Populations.* New York, Dover Press.
5. Cavalli-Sforza LL, Feldman MW 2003. Biology as history: population genetic approaches to modern human evolution. *Nature Genetics* 33:266-275
6. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994 *The History and Geography of Human Genes.* Princeton NJ, Princeton Univ.Press.
7. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J 1988. Reconstruction of human evolution: bringing together genetic, archeological and linguistic data. *Proc. Natl. Acad. Sci. USA* 85:6002-6.
8. Dupanloup I, Pereira L, Bertorelle G, Calafell F, João Prata M, Amorim A, Barbujani G. 2003. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *Journal of Molecular Evolution.* 57:85-97.
9. Fisher RA 1930. *The Genetical Theory of Natural Selection.* New York, Dover.
10. Fisher RA 1937 The wave of advance of advantageous genes. *Ann.Eugen.*7:355-69
11. Kimura M 1968. Evolutionary rate at the molecular level. *Nature* 217:624-6
12. Liu H, Prugnolle F, Manica A, Balloux F. A geographically explicit model of worldwide Human-Settlement theory. *Am. J. Hum. Genet.* 79:230-7.
13. Ohta T, Kimura M, 1970. Development of associative overdominance through linkage disequilibrium in finite populations. *Genet. Res.* 16:165-77.
14. Prugnolle F, Manica A, Balloux F, 2005. Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* 15:1022-7
15. Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman, MW, Cavalli-Sforza LL 2005 Support for a serial founder effect originating in Africa, using the relationship between genetic and geographic distance in human populations. *Proc. Natl. Acad. Sci. USA* 102:15492-7
16. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd HM, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* 298:2381-5.