GIORGIO BERNARDI

La teoria neoselezionista dell'evoluzione

In biologia esistono problemi di dettaglio e problemi di natura fondamentale, ossia problemi la cui soluzione porta a conclusioni di carattere generale. I primi sono numerosissimi, i secondi molto pochi. Uno di questi, forse il più importante (tanto che Jacques Monod ne ha fatto il tema centrale del suo famoso libro del 1970 *Le hasard et la nécessité. Essai sur la philosophie naturelle de la biologie moderne*), riguarda il ruolo del caso nell'evoluzione degli organismi viventi.

In realtà, ogni teoria dell'evoluzione, dalle teorie classiche alla teoria neoselezionista, ha portato a una visione particolare di questo problema, in quanto ha dato una valutazione, seppure necessariamente qualitativa, delle mutazioni neutrali (che possono essere prese come un'indicazione della parte del caso nell'evoluzione; v. oltre). Qui di seguito saranno presentate in modo sintetico le valutazioni proposte dalle teorie classiche dell'evoluzione per poi affrontare il soggetto principale di questo saggio, la teoria neoselezionista. Premettiamo che qui accetteremo come definizione di teoria scientifica ogni tentativo di collegare in modo sistematico le conoscenze riguardanti un determinato aspetto del mondo dell'esperienza (M. Ruse, *Theory*, in *The Oxford companion to philosophy*, ed. T. Honderich, 1995). Secondo questa visione classica, una teoria è un sistema *ipotetico-deduttivo*, cioè un insieme di leggi che sono la conseguenza deduttiva di poche solide ipotesi.

Le teorie classiche dell'evoluzione

La frase più famosa nel libro *On the origin of species* (1859) di Charles Darwin è certamente quella che definisce l'essenza stessa della teoria, cioè la selezione naturale: «Chiamo Selezione Naturale la conservazione delle variazioni favorevoli e l'eliminazione delle variazioni dannose». Questa definizione suggerisce una dicotomia per quanto riguarda il destino delle «variazioni», cioè delle mutazioni, per usare un linguaggio moderno, ed è stata generalmente interpretata in questo senso. Tuttavia, questa frase è immediatamente seguita da un'altra che non è quasi mai citata: «Le variazioni che non sono né utili né dannose non sono soggette alla selezione naturale». In altre parole, Darwin distinse non due, ma tre specie di variazioni: vantaggiose, deleterie e neutre. Mentre le variazioni vantaggiose si propagano nella progenie per selezione positiva, le variazioni deleterie tendono a scomparire per selezione negativa. Nel primo caso, i discendenti dei portatori delle mutazioni aumenteranno nella popolazione considerata, perché si riprodurranno di più, nel secondo diminuiranno, perché si riprodurranno di meno. Le variazioni neutre possono uscire dal loro limbo per essere fissate come le variazioni vantaggiose, o per sparire come quelle deleterie. È interessante notare che il concetto di variazioni neutre (assente in Alfred Russel Wallace) è un'importante scoperta di Darwin, anche se generalmente non apprezzata come tale. Poiché, almeno in prima approssimazione, le mutazioni sono eventi aleatori, e poiché le mutazioni neutre rappresentano l'enorme maggioranza delle mutazioni (almeno nei genomi qui considerati), la loro valutazione permette di stimare la parte del caso nell'evoluzione.

Le variazioni neutre furono semplicemente ignorate dai neodarwinisti Ronald Aylmer Fisher e John B.S. Haldane. Esse risorsero, tuttavia, grazie alla teoria neutralista, che ruppe il lungo predominio delle teorie selezioniste dei neodarwinisti. Il promotore della teoria neutralista, Motoo Kimura (1924-1994), propose che «la causa principale dei cambiamenti nell'evoluzione a livello molecolare», ossia nel DNA o acido desossiribonucleico «è la fissazione a caso di mutanti selettivamente neutri o quasi neutri». La «sopravvivenza del più adatto» di Darwin era in questo modo sostituita dalla «sopravvivenza del più fortunato» di Kimura, e l'evoluzione darwiniana e neodarwiniana dall'«evoluzione non darwiniana» (J.L. King, T.H. Jukes, *Non-Darwinian evolution*, «Science», 1969, 3881, pp. 788-98). La proposta rivoluzionaria di Kimura lanciò un dibattito tra neutralisti e selezionisti che è

ancora in corso, sotto nuove fórme, quarant'anni dopo il suo inizio. Una modificazione significativa della teoria neutralista è stata la *teoria quasi neutralista* di Tomoko Ohta (2002). Questa teoria propone che una parte sostanziale dei cambiamenti sia provocata dalla fissazione aleatoria di cambiamenti *quasi neutri*, cioè da una classe di cambiamenti che «include cambiamenti intermedi tra neutri e vantaggiosi come anche tra neutri e deleteri». I primi quattro schemi della figura 1 riassumono graficamente i punti appena esposti.

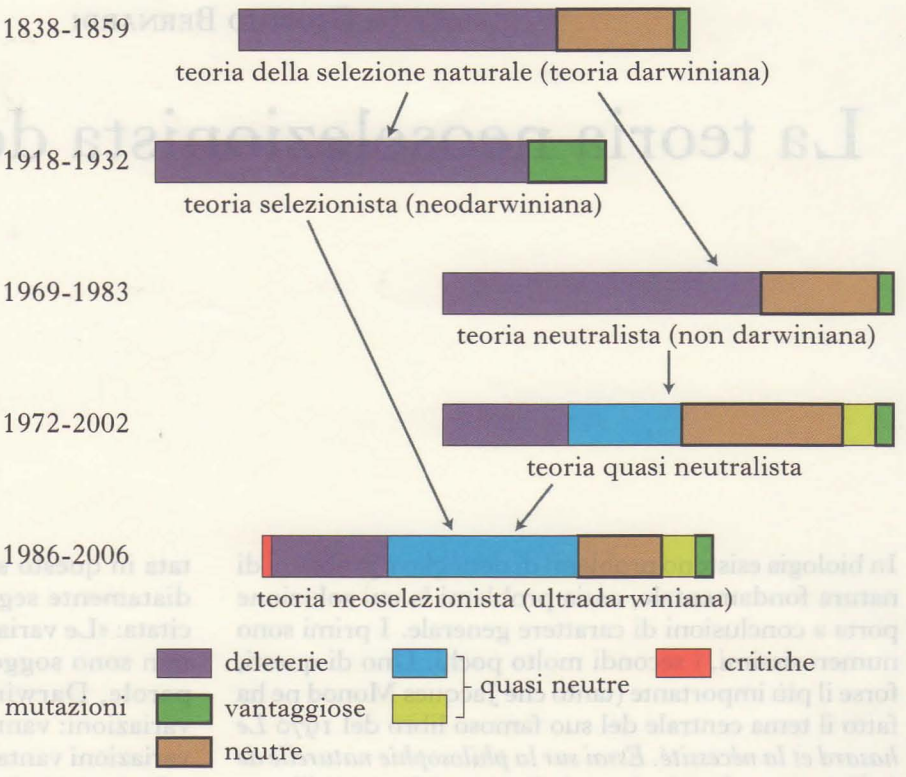

Fig. 1 – Teorie dell'evoluzione

Sono indicati i periodi del loro sviluppo e il peso dato ai vari tipi di mutazione nel genoma (Fonte: figura ridisegnata da Bernardi 2007)

Gli approcci sperimentali delle teorie classiche

Questi approcci devono essere considerati con grande attenzione poiché esistono legami stretti tra approcci sperimentali, risultati e conclusioni generali. La selezione naturale agisce sul *fenotipo*, cioè sui caratteri osservabili negli organismi viventi (forme, funzioni, proprietà in generale). È quindi perfettamente comprensibile che il primo approccio allo studio dell'evoluzione si sia basato su caratteri morfologici. Il caso classico è quello dei becchi dei fringuelli delle isole Galápagos, i quali mostrano adattamenti a diversi tipi di alimenti che possono andare da semi duri a tessuti vegetali molli. Oggi sappiamo che la morfologia del becco dei fringuelli di Darwin dipende da una dozzina di geni. Dopo la riscoperta delle leggi di Gregor Mendel all'inizio del 20° sec., i neodarwinisti si basarono sullo studio della trasmissione di un piccolo numero di caratteri genetici da una generazione all'altra. Solo più tardi si sviluppò un approccio molecolare sulla base delle prime sequenze disponibili di proteine e di geni. Fino a questo punto, tutti gli approcci descritti sono stati basati su quello che è stato chiamato *fenotipo classico* (morfologico, genetico, o molecolare), determinato, in ultima analisi, da un piccolo numero di geni. L'approccio composizionale da noi sviluppato si fonda invece sul *fenotipo del genoma*, cioè sulle caratteristiche composizionali dell'intero genoma. La descrizione di questo approccio richiede, tuttavia, una breve premessa che riguarda la biologia molecolare e la genomica.

La biologia molecolare

La biologia molecolare ha posto le basi della biologia moderna intorno alla metà del 20° sec., semplificando e chiarendo la nostra visione della complessità degli organismi viventi con lo stabilire tre nozioni fondamentali: 1) *la struttura a doppia elica del DNA* (James Dewey Watson e Francis Harry Compton Crick; fig. 2*A*); in questa struttura, formata da due catene, le eliche, antiparallele (cioè con direzione opposta) di polinucleotidi, le quattro basi presenti nei nucleotidi sono accoppiate nelle forma G (guanina) con C (citosina), A (adenina) con T (timina); questa struttura (che spiega, per inciso, come la composizione del DNA possa essere data dalla percentuale di G+C, o GC) ha permesso di comprendere il meccanismo semiconservativo della replicazione del DNA (Matthew Stanley Meselson e Franklin William Stahl), ciascuna elica essendo copiata da una nuova elica complementare della prima con una grandissima fedeltà (gli errori nella copia corrispondono alle mutazioni, in cui una base è copiata non nella base complementare, ma in un'altra base); 2) *il codice genetico* (Marshall Warren Nirenberg e Johann Heinrich Matthaei; fig. 2*B*), cioè la correlazione tra i codoni del DNA e gli amminoacidi delle proteine; i codoni sono costituiti da triplette di nucleotidi e sono tradotti in amminoacidi secondo un preciso codice: per es., ATG codifica la metionina, TCG il triptofano ecc.; tre codoni corrispondono ad altrettanti segnali di terminazione; 3) *la regolazione dell'espressione* dei geni in proteine (François Jacob e Jacques Monod; fig. 2*C*); questa, la più complessa

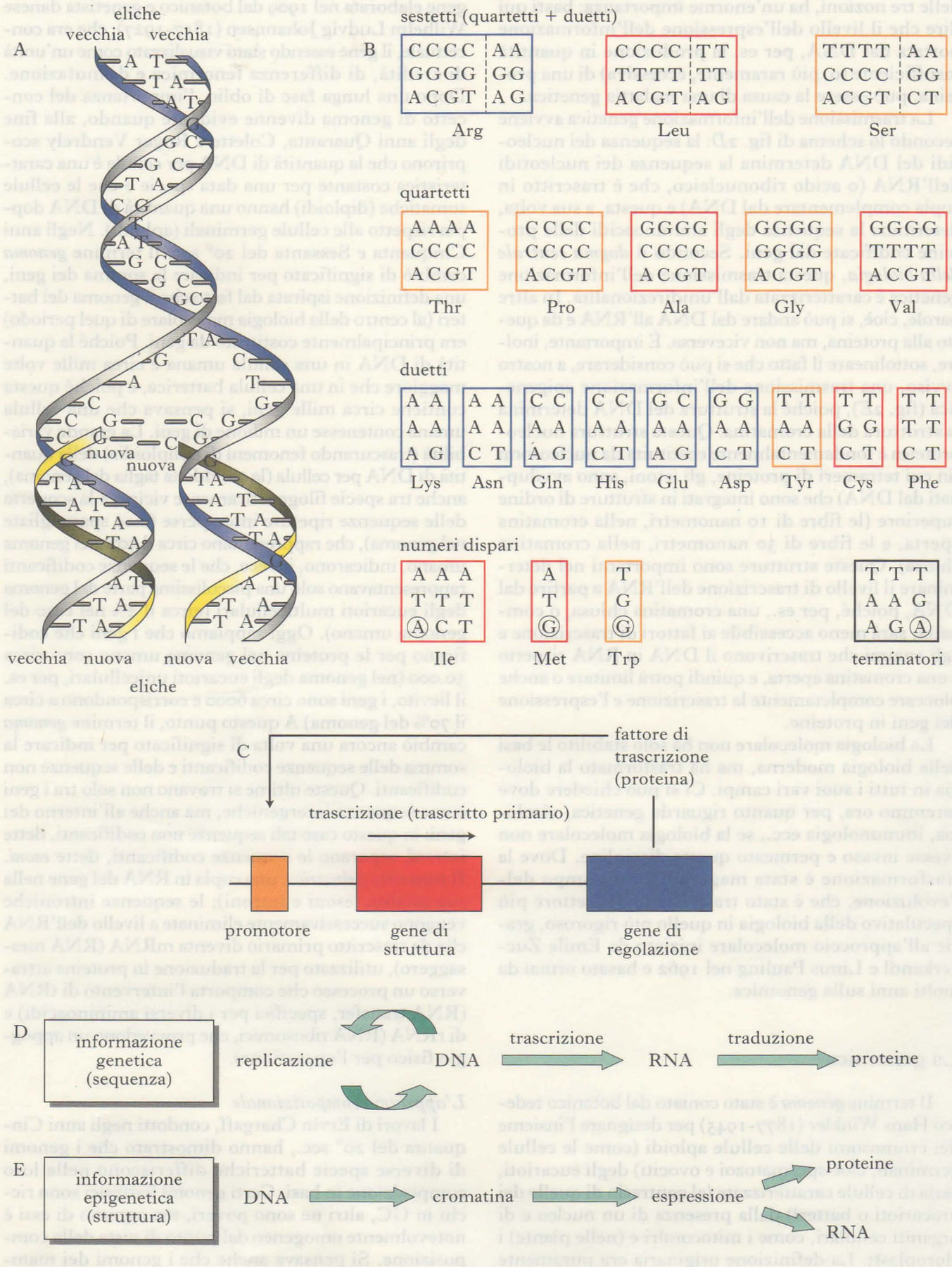

Fig. 2 – Le caratteristiche del DNA

delle tre nozioni, ha un'enorme importanza; basti qui dire che il livello dell'espressione dell'informazione portata dal DNA, per es. la produzione in quantità insufficiente (o, più raramente, eccessiva) di una proteina, può essere la causa di una malattia genetica.

La trasmissione dell'informazione genetica avviene secondo lo schema di fig. 2*D*: la sequenza dei nucleotidi del DNA determina la sequenza dei nucleotidi dell'RNA (o acido ribonucleico, che è trascritto in copia complementare dal DNA) e questa, a sua volta, determina la sequenza degli amminoacidi delle proteine codificate dai geni. Secondo il *dogma centrale della biologia*, questa trasmissione dell'informazione genetica è caratterizzata dall'unidirezionalità. In altre parole, cioè, si può andare dal DNA all'RNA e da questo alla proteina, ma non viceversa. È importante, inoltre, sottolineare il fatto che si può considerare, a nostro avviso, una trasmissione dell'informazione epigenetica (fig. 2*E*), poiché la struttura del DNA determina la struttura della cromatina. Questa struttura nucleoproteica è fondamentalmente costituita da nucleosomi (in cui tetrameri di proteine, gli istoni, sono avvilupati dal DNA) che sono integrati in strutture di ordine superiore (le fibre di 10 nanometri, nella cromatina aperta, e le fibre di 30 nanometri, nella cromatina chiusa). Queste strutture sono importanti nel determinare il livello di trascrizione dell'RNA a partire dal DNA, poiché, per es., una cromatina chiusa, o compatta, sarà meno accessibile ai fattori di trascrizione e agli enzimi che trascrivono il DNA in RNA rispetto a una cromatina aperta, e quindi potrà limitare o anche bloccare completamente la trascrizione e l'espressione dei geni in proteine.

La biologia molecolare non ha solo stabilito le basi della biologia moderna, ma ha trasformato la biologia in tutti i suoi vari campi. Ci si può chiedere dove saremmo ora, per quanto riguarda genetica, virologia, immunologia ecc., se la biologia molecolare non avesse invaso e permeato queste discipline. Dove la trasformazione è stata maggiore è nel campo dell'evoluzione, che è stato trasformato dal settore più speculativo della biologia in quello più rigoroso, grazie all'approccio molecolare iniziato da Emile Zuckerkandl e Linus Pauling nel 1962 e basato ormai da molti anni sulla genomica.

La genomica

Il termine *genoma* è stato coniato dal botanico tedesco Hans Winkler (1877-1945) per designare l'insieme dei cromosomi delle cellule aploidi (come le cellule germinali, cioè spermatozoi e ovociti) degli eucarioti, ossia di cellule caratterizzate (al contrario di quelle dei procarioti o batteri) dalla presenza di un nucleo e di organiti cellulari, come i mitocondri e (nelle piante) i cloroplasti. La definizione originaria era puramente operativa, a differenza della più antica definizione di gene elaborata nel 1909 dal botanico e genetista danese Wilhelm Ludvig Johannsen (1857-1927), che era concettuale, il gene essendo stato visualizzato come un'unità di eredità, di differenza fenotipica e di mutazione. Dopo una lunga fase di oblio, l'importanza del concetto di genoma divenne evidente quando, alla fine degli anni Quaranta, Colette e Roger Vendrely scoprirono che la quantità di DNA per cellula è una caratteristica costante per una data specie e che le cellule somatiche (diploidi) hanno una quantità di DNA doppia rispetto alle cellule germinali (aploidi). Negli anni Cinquanta e Sessanta del 20° sec. il termine *genoma* cambiò di significato per indicare la somma dei geni, una definizione ispirata dal fatto che il genoma dei batteri (al centro della biologia molecolare di quel periodo) era principalmente costituito da geni. Poiché la quantità di DNA in una cellula umana è circa mille volte maggiore che in una cellula batterica, e poiché questa contiene circa mille geni, si pensava che una cellula umana contenesse un milione di geni. La grande variabilità (trascurando fenomeni di poliploidia) della quantità di DNA per cellula (la cosiddetta taglia del genoma), anche tra specie filogeneticamente vicine, e la scoperta delle sequenze ripetute intersperse (cioè sparpagliate nel genoma), che rappresentano circa il 50% del genoma umano, indicarono, invece, che le sequenze codificanti rappresentavano solo una piccolissima parte del genoma degli eucarioti multicellulari (circa il 2% nel caso del genoma umano). Oggi sappiamo che i geni che codificano per le proteine nel genoma umano sono circa 30.000 (nel genoma degli eucarioti unicellulari, per es. il lievito, i geni sono circa 6000 e corrispondono a circa il 70% del genoma) A questo punto, il termine *genoma* cambiò ancora una volta di significato per indicare la somma delle sequenze codificanti e delle sequenze non codificanti. Queste ultime si trovano non solo tra i geni come sequenze intergeniche, ma anche all'interno dei geni: in questo caso tali sequenze non codificanti, dette *introni*, separano le sequenze codificanti, dette *esoni*. Il *trascritto primario* è una copia in RNA del gene nella sua totalità (esoni e introni); le sequenze introniche vengono successivamente eliminate a livello dell'RNA che da trascritto primario diventa mRNA (RNA messaggero), utilizzato per la traduzione in proteina attraverso un processo che comporta l'intervento di tRNA (RNA transfer, specifici per i diversi amminoacidi) e di rRNA (RNA ribosomici, che provvedono un appoggio fisico per l'operazione).

L'approccio composizionale

I lavori di Ervin Chargaff, condotti negli anni Cinquanta del 20° sec., hanno dimostrato che i genomi di diverse specie batteriche differiscono nella loro composizione in basi. Certi genomi batterici sono ricchi in GC, altri ne sono poveri, ma ognuno di essi è notevolmente omogeneo dal punto di vista della composizione. Si pensava anche che i genomi dei mammiferi (come il genoma bovino) fossero molto più ete-

rogenei in composizione dei genomi batterici (anche se questa idea era dovuta alla presenza di DNA satelliti, costituiti da lunghissimi tratti di corte sequenze ripetute presenti nei centromeri dei cromosomi). Il punto di partenza del nostro approccio composizionale, iniziato cinquant'anni fa, è consistito nel degradare il DNA di mammiferi e uccelli con una DNasi (cioè un enzima che taglia il DNA) e nel frazionare il DNA così tagliato su colonne di un fosfato di calcio, l'idrossiapatite, una tecnica cromatografica che avevamo appena sviluppato. Si è trattato del primo tentativo mirato ad analizzare il genoma degli eucarioti o, se si vuole, il primo esperimento di una nuova disciplina che si sarebbe chiamata, molti anni dopo, *genomica*. La risoluzione ottenuta su idrossiapatite non è stata però sufficiente, salvo in casi particolari quali, per es., il genoma mitocondriale del lievito, in cui una buona parte del genoma è costituita da sequenze intergeniche quasi esclusivamente formate da A e T. La maggior parte dei nostri lavori successivi si è basata, perciò, sul frazionamento del DNA per ultracentrifugazione in un gradiente di densità di solfato di cesio in presenza di ligandi (quali lo ione argento) specifici per corte sequenze di nucleotidi del DNA. Questo approccio ha permesso di frazionare i complessi DNA-ligando sulla base della densità del ligando sul DNA. Per es., se il ligando si lega a sequenze ricche in GC, le molecole di DNA ricche in GC sono appesantite e si separano dalle molecole povere in GC. Tale approccio composizionale è stato in seguito facilmente spostato dall'analisi del DNA dopo ultracentrifugazione in gradienti di densità alle sequenze nucleotidiche, appena queste sono divenute disponibili grazie alle tecniche di sequenziamento di interi genomi.

Malgrado la sua complessità sperimentale, l'approccio composizionale è caratterizzato da una grande semplicità concettuale. Infatti esso si basa sulla proprietà più elementare del DNA, la sua composizione. Questo approccio si presta a studiare sia la struttura sia l'evoluzione del genoma eucariotico, poiché la composizione in basi non solo influenza la struttura del DNA e, di conseguenza, la struttura della cromatina, ma può inoltre essere alterata da mutazioni, inserzioni e delezioni, come anche da ricombinazioni e traslocazioni. Esso inoltre presenta due vantaggi rispetto agli approcci precedenti: da un lato la grande precisione, dall'altro l'estensione dell'analisi da pochi geni all'insieme del genoma (o a sue regioni). Inoltre, la strategia dell'approccio composizionale è molto diversa dalle strategie precedenti in quanto si parte dall'insieme per scendere al particolare e non viceversa. Questa strategia ha permesso quindi di vedere gli aspetti generali per primi, per poi discendere nei dettagli. Nella strategia opposta occorre, invece, conoscere molti dettagli prima di concludere che essi riflettono aspetti generali. Esamineremo ora i risultati ottenuti grazie all'approccio composizionale, prima a livello della struttura e poi dell'evoluzione del genoma.

La struttura del genoma eucariotico

L'approccio composizionale ha condotto a tre scoperte importanti che ci hanno permesso di definire alcune proprietà generali del genoma degli eucarioti. 1) I genomi dei vertebrati (e degli eucarioti in generale) sono *mosaici di isocore*, ossia regioni che contengono da centinaia di migliaia a milioni di paia di basi. Le isocore hanno una composizione piuttosto uniforme (fig. 3*A*) e appartengono a un piccolo numero di famiglie (L1, L2, H1, H2 e H3) caratterizzate da differenti livelli di GC e da diverse quantità relative (fig. 4). Le famiglie di isocore di un genoma rappresentano un fenotipo, che abbiamo chiamato *fenotipo del genoma*. La composizione in basi delle famiglie di isocore del genoma umano copre uno spettro composizionale che va dal 30 al 60% di GC, uno spettro paragonabile a

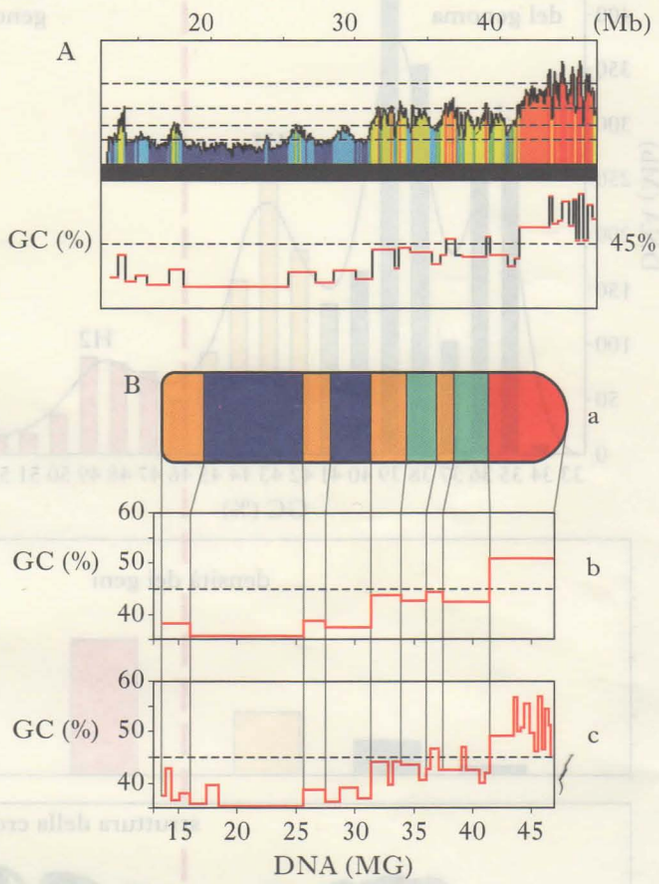

Fig. 3 – Profilo composizionale del cromosoma 21 umano

In *A*, il diagramma superiore rappresenta il profilo di GC in segmenti successivi, ognuno di 100 kb, e i colori da blu a rosso corrispondono a livelli crescenti di GC. Le linee orizzontali rosse nel diagramma inferiore indicano le isocore. In *B*, *a*, ideogramma ad alta risoluzione (850 bande); *b*, profilo di GC; *c*, profilo delle isocore (Fonti: *A*, figura ridisegnata da Costantini, Clay, Auletta, Bernardi 2006; *B*, figura ridisegnata da Costantini, Clay, Federico et al. 2007)

quello presentato dall'insieme dei genomi batterici. È importante sottolineare il carattere discontinuo, a mosaico, della composizione in basi del genoma, poiché l'idea prevalente fino ai nostri lavori era che la variazione della composizione del genoma fosse continua. Ricordiamo che il genoma umano comprende 3200 milioni di paia di basi; le isocore hanno dimensioni comprese tra 0,5 e 1 milione di paia di basi.

2) In questo *paesaggio composizionale* del genoma (per es., il genoma umano) le regioni ricche (più del 46%) in GC (che noi abbiamo chiamato *cuore del genoma* per ragioni che diventeranno presto evidenti) rappresentano il 15% del genoma, mentre le regioni povere (meno del 46%) in GC (*deserto del genoma*) corrispondono al rimanente 85%. Il cuore del genoma non è solo caratterizzato dalla sua ricchezza in GC, ma anche da tutta una serie di proprietà strutturali e funzionali molto importanti (fig. 4): un'alta concentrazione in geni, una struttura aperta della cromatina (il che spiega la sua predisposizione ad accettare delezioni e inserzioni, come l'integrazione di sequenze virali), una bassa metilazione del DNA, un'alta concentrazione in sequenze ripetute intersperse corte o SINEs (*Short INterspersed Elements*, ossia sequenze molto simili tra loro, lunghe circa 300 paia di basi e presenti in circa 1,2 milioni di copie nel genoma umano), ma una bassa concentrazione in sequenze ripetute lunghe o LINEs (*Long INterspersed Elements*, che arrivano fino a 6000 paia di basi); inoltre le velocità di mutazione e di ricombinazione sono elevate, la replicazione è precoce nel ciclo cellulare, l'espressione dei geni è elevata. Il deserto del genoma è caratterizzato da proprietà opposte a quelle appena menzionate. Dobbiamo anche ricordare che le isocore sono le bande cromosomiche a più alta risoluzione (fig. 3*B*), sono unità di replicazione e sono caratterizzate da percentuali di dinucleotidi (per es. AA, GC, CG ecc.), di trinucleotidi (AAC, CAT ecc.) e di codoni (i trinucleotidi che codificano gli amminoacidi; v. fig. 2) correlate ai livelli di GC delle isocore (per inciso, questo spiega il successo dell'approccio sperimentale basato sull'ultracentrifugazione di complessi DNA/ligandi).

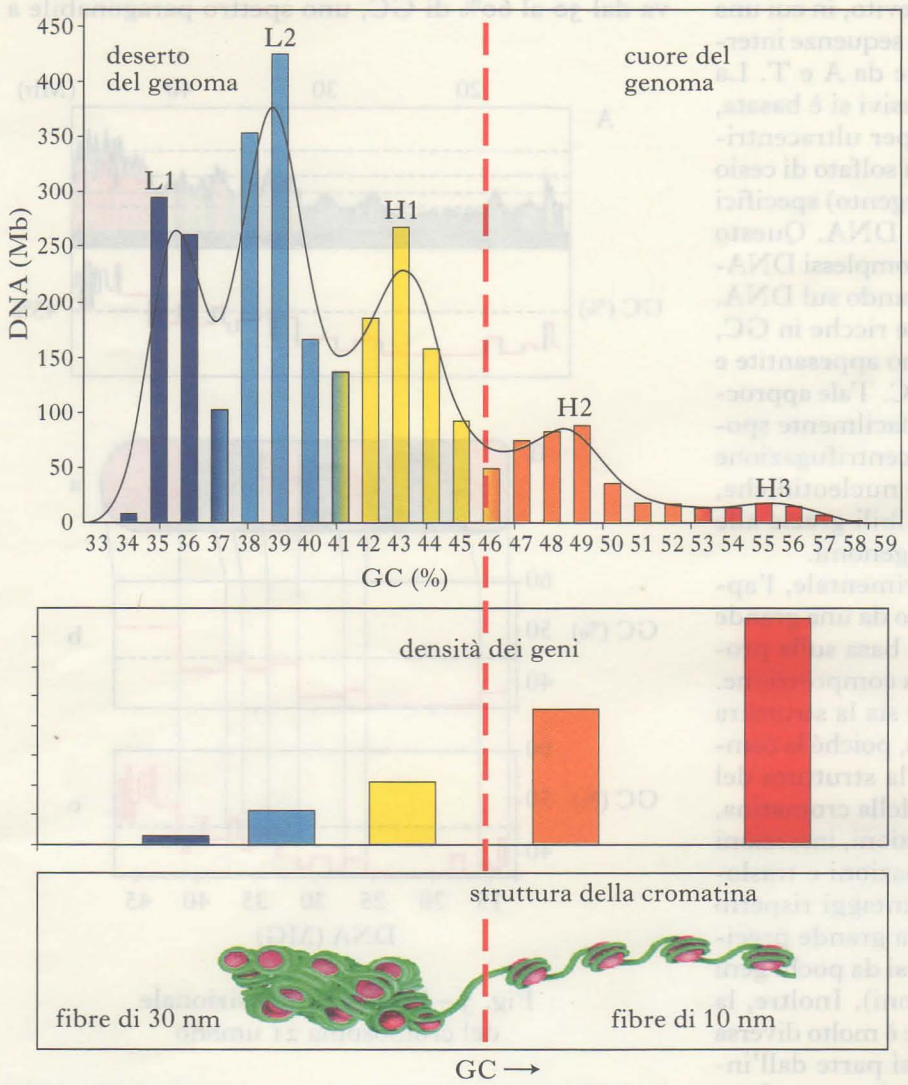

Fig. 4 – Proprietà fondamentali del cuore e del deserto del genoma

In alto famiglie di isocore, al centro loro densità in geni e in basso struttura della cromatina. A sinistra, nella cromatina compatta, i nucleosomi sono addossati tra loro; a destra, nella cromatina aperta, sono spaziati (Fonti: in alto, figura ridisegnata da Costantini, Clay, Auletta, Bernardi 2006; in basso e al centro, figure ridisegnate da Bernardi 2007)

Un punto importante che appare sempre più chiaramente è che il deserto del genoma, lungi dall'essere formato da grandi regioni inerti, comprende soprattutto geni implicati nello sviluppo dell'organismo, cioè geni di grandi dimensioni, con sequenze di regolazione molto complesse, costituite da diversi moduli spesso molto lontani dai geni interessati. Questo è comprensibile visto che tali sequenze devono determinare non solo le quantità di proteine prodotte, ma anche la fase di sviluppo in cui l'espressione deve avvenire. Inoltre, questo spiega come alla fine dello sviluppo il deserto del genoma sia bloccato da una struttura chiusa della cromatina. Invece, il cuore del genoma comprende principalmente geni di mantenimento della vita cellulare (geni *housekeeping*), che richiedono una regolazione molto più semplice. Il

fatto che i sistemi di regolazione siano diversi nelle diverse isocore suggerisce che queste possano essere responsabili per processi di coregolazione di proteine correlate dal punto di vista funzionale.

3) Un *codice genomico* (da non confondere con il codice genetico, che stabilisce la corrispondenza dei codoni con gli amminoacidi) correla le composizioni: delle sequenze codificanti e delle sequenze non codificanti contigue (cioè del 2% del genoma con il restante 98%, nel caso del genoma dei mammiferi); delle sequenze codificanti con la composizione in amminoacidi e con la struttura secondaria delle proteine; della prima, seconda e terza posizione dei codoni tra di loro. Inoltre, come già detto, la struttura del DNA determina la struttura della cromatina. Ciò ha condotto alla conclusione che le isocore sono una struttura fondamentale del genoma degli eucarioti e che questo è un *insieme integrato*. La correlazione composizionale tra sequenze codificanti e non codificanti (intergeniche e introniche) indica l'esistenza di contesti genomici. Tali contesti sono importanti per l'espressione dei geni, come è dimostrato dal fatto che una sequenza retrovirale, se inserita in un contesto composizionale correlato (ossia rispettoso delle correlazioni dei geni dell'ospite con le sequenze non codificanti contigue), sarà stabile e si esprimerà. Se inserita in un contesto composizionale non correlato (cioè molto diverso) non sarà stabile e non sarà espressa. È importante sottolineare che le sequenze non codificanti, rappresentanti l'enorme maggioranza del genoma dei vertebrati, sono soggette a costrizioni composizionali, cioè non sono libere di variare in composizione, ma variano insieme con le sequenze codificanti contigue.

L'evoluzione composizionale del genoma

La nuova visione del genoma appena descritta ha delle implicazioni importanti per quanto riguarda la sua evoluzione. Infatti, nessun processo aleatorio, quali, per es., le mutazioni puntiformi (cioè i cambiamenti di singole paia di basi) può: 1) condurre alle correlazioni composizionali (il codice genomico menzionato più sopra); 2) permettere una correlazione tra le isocore e le proprietà strutturali e funzionali del genoma; 3) creare e/o mantenere un mosaico di isocore con frontiere ben delimitate e conservate nell'evoluzione. Queste scoperte escludono il modello di un genoma in cui i geni sono distribuiti a caso nella massa delle sequenze non codificanti o, anche, di un genoma dotato di sole proprietà additive (in cui i geni e le loro sequenze regolatrici sono cassette che funzionano ugualmente qualunque sia la loro posizione nel genoma) e non di proprietà cooperative (come quelle dettate dal contesto genomico; v. *supra*). Inoltre, il semplice confronto dei nostri risultati iniziali basati sull'ultracentrifugazione (confermati da nostri dati successivi basati sulle sequenze

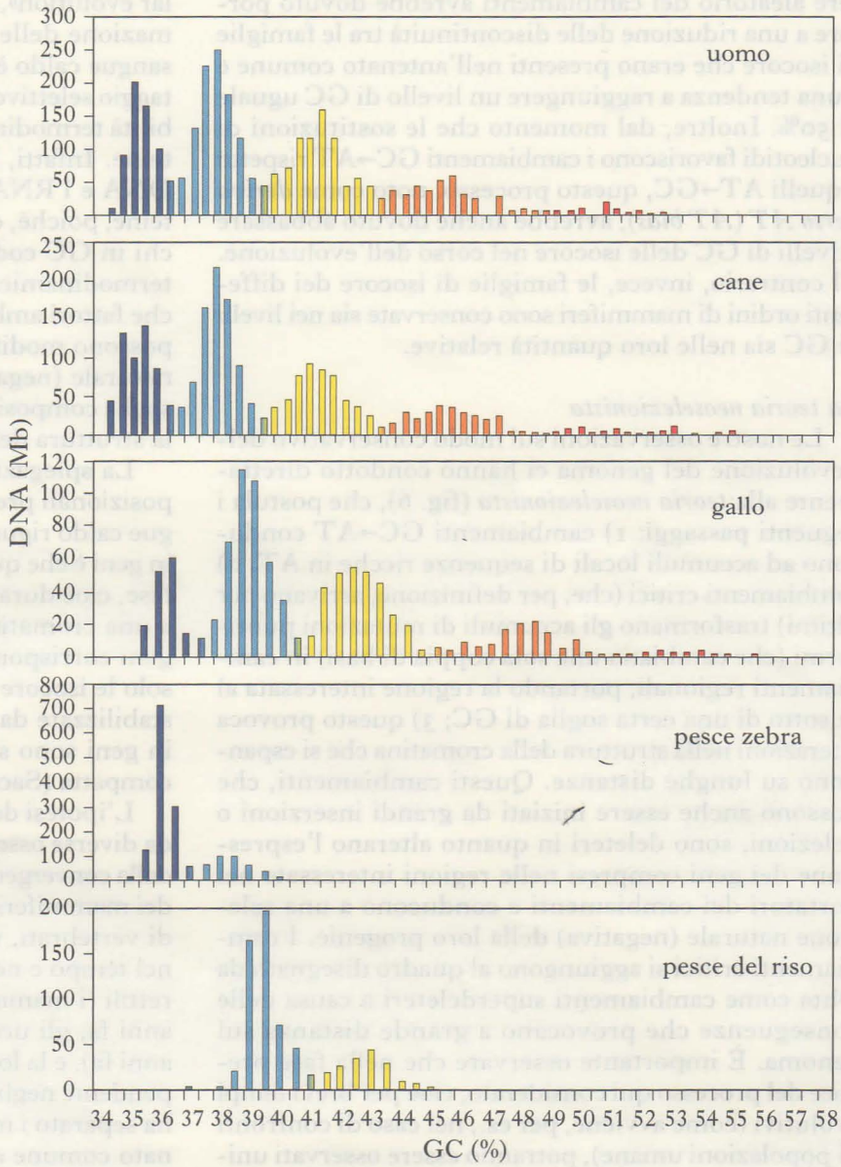

Fig. 5 – Famiglie di isocore in alcuni genomi di vertebrati

(Fonte: figura ridisegnata da Costantini, Auletta, Bernardi 2007 e Costantini, Di Filippo, Auletta, Bernardi 2007)

dei genomi dei vertebrati) ci ha portato alla scoperta di due tipi di evoluzione composizionale del genoma: il *modo conservativo* e il *modo transitivo*.

Il modo conservativo

Questo modo dell'evoluzione del genoma è esemplificato dall'enorme somiglianza delle famiglie di isocore dei genomi dei primati e dei carnivori (fig. 5), nonché dalla grandissima somiglianza composizionale di regioni corrispondenti dei loro cromosomi. Almeno la metà delle coppie di basi di questi genomi è cambiata nei circa 90 milioni di anni che separano l'antenato comune da questi due ordini di mammiferi, i quali ne sono derivati indipendentemente l'uno dall'altro (cioè senza scambi genetici tra loro). Il carattere aleatorio dei cambiamenti avrebbe dovuto portare a una riduzione delle discontinuità tra le famiglie di isocore che erano presenti nell'antenato comune e a una tendenza a raggiungere un livello di GC uguale al 50%. Inoltre, dal momento che le sostituzioni di nucleotidi favoriscono i cambiamenti GC→AT rispetto a quelli AT→GC, questo processo, noto come *deriva verso AT* (*AT bias*), avrebbe anche dovuto abbassare i livelli di GC delle isocore nel corso dell'evoluzione. Al contrario, invece, le famiglie di isocore dei differenti ordini di mammiferi sono conservate sia nei livelli di GC sia nelle loro quantità relative.

La teoria neoselezionista

Le nostre osservazioni sul modo conservativo dell'evoluzione del genoma ci hanno condotto direttamente alla *teoria neoselezionista* (fig. 6), che postula i seguenti passaggi: 1) cambiamenti GC→AT conducono ad accumuli locali di sequenze ricche in AT; 2) cambiamenti critici (che, per definizione, arrivano per ultimi) trasformano gli accumuli di mutazioni puntiformi (che cambiano una sola coppia di basi) in cambiamenti regionali, portando la regione interessata al di sotto di una certa soglia di GC; 3) questo provoca alterazioni nella struttura della cromatina che si espandono su lunghe distanze. Questi cambiamenti, che possono anche essere iniziati da grandi inserzioni o delezioni, sono deleteri in quanto alterano l'espressione dei geni compresi nelle regioni interessate nei portatori dei cambiamenti e conducono a una selezione naturale (negativa) della loro progenie. I cambiamenti critici si aggiungono al quadro disegnato da Ohta come cambiamenti superdeleteri a causa delle conseguenze che provocano a grande distanza sul genoma. È importante osservare che nella fase precoce del processo qui considerato, cioè per brevi tempi evolutivi (come avviene, per es., nel caso di confronti di popolazioni umane), potranno essere osservati unicamente cambiamenti in maggioranza neutrali o quasi neutrali, mentre su tempi lunghi i fenomeni di selezione saranno ben visibili, in quanto saranno sopravvissuti solo gli individui che non sono stati eliminati dalla selezione negativa.

Il modo transitivo

Questo modo dell'evoluzione del genoma è stato scoperto grazie alla nostra osservazione che i genomi di pesci, anfibi e molti rettili presentano isocore (in rosa in fig. 7) differenti da quelle ricche in GC caratterizzanti i genomi dei vertebrati a sangue caldo (in rosso in fig. 7), in quanto le isocore ricche in geni dei vertebrati a sangue freddo sono solo moderatamente più ricche in GC del resto del genoma. Il deserto del genoma (in blu in fig. 7) non subisce cambiamenti composizionali importanti. È quindi evidente che deve esistere anche un modo transitivo nell'evoluzione del genoma. Una spiegazione fornita più di venti anni fa (Giorgio Bernardi, Giacomo Bernardi, *Compositional constraints and genome evolution*, «Journal of molecular evolution», 1986, 1 e 2, pp. 1-11) è stata che la formazione delle isocore ricche in GC dei vertebrati a sangue caldo è dovuta alla selezione naturale. Il vantaggio selettivo è stato individuato nell'aumentata stabilità termodinamica del DNA, dell'RNA e delle proteine. Infatti, l'aumento di GC stabilizza non solo il DNA e l'RNA (che ne è la copia), ma anche le proteine, poiché, come abbiamo dimostrato, i codoni ricchi in GC codificano amminoacidi che stabilizzano termodinamicamente le proteine. È evidente quindi che fattori ambientali (nel nostro caso la temperatura) possono modificare il genoma attraverso la selezione naturale (negativa), innalzando progressivamente la soglia composizionale al di sotto della quale si alterano la struttura della cromatina e l'espressione dei geni.

La spiegazione del fatto che i cambiamenti composizionali presentati dai genomi dei vertebrati a sangue caldo riguardano essenzialmente le isocore ricche in geni è che queste corrispondono (nei nuclei in interfase, cioè durante la maggior parte del ciclo cellulare) a una cromatina aperta, mentre le isocore povere in geni corrispondono a una cromatina chiusa. Quindi solo le isocore ricche in geni hanno bisogno di essere stabilizzate da aumenti di GC, mentre quelle povere in geni sono stabilizzate dalla loro stessa cromatina compatta (Saccone, Federico, Bernardi 2002).

L'ipotesi della stabilità termodinamica è sostenuta da diverse osservazioni. Una di queste è rappresentata dalla convergenza dei profili composizionali dei genomi dei mammiferi e degli uccelli (fig. 5). Queste due classi di vertebrati, infatti, hanno avuto un'origine diversa nel tempo e nella derivazione da due diverse classi di rettili (i mammiferi dai terapsidi circa 220 milioni di anni fa, gli uccelli dai dinosauri circa 150 milioni di anni fa), e la loro evoluzione è avvenuta in modo indipendente negli ultimi 300 milioni di anni (il tempo che ha separato i mammiferi e gli uccelli attuali dall'antenato comune a tutti gli amnioti, cioè rettili, uccelli e mammiferi). Questo fa pensare che le stesse cause (in questo caso l'omeotermia a 37-41 °C) hanno provocato gli stessi effetti. Una seconda osservazione riguarda il predominio dei cambiamenti AT→GC su quelli GC→AT che si evidenziano nei geni di *Gillichtys seta*,

un pesce che vive a 40-41 °C, rispetto agli stessi geni di *Gillichtys mirabilis* che vive a 15-20 °C (Bucciarelli, Di Filippo, Costagliola et al. 2009). Le regioni ricche in geni del genoma del pesce che vive ad alta temperatura presentano inoltre un'amplificazione di sequenze intersperse ricche in GC che stabilizzano queste regioni (tale risultato suggerisce una funzione analoga per le sequenze ripetute SINEs, ricche in GC, che sono abbondanti nel cuore del genoma dei vertebrati a sangue caldo). Infine, sono stati osservati cambiamenti AT→GC in RNA ribosomiali di vertebrati che vivono a temperature più alte anche di pochi gradi rispetto ai vertebrati di confronto.

Mentre l'innalzamento della temperatura corporea (quale che sia la sua origine, ossia omeotermia, temperatura ambientale ecc.) è certamente il *primum movens* delle transizioni composizionali del genoma che accompagnarono l'emergenza dei mammiferi e degli uccelli, altri fattori quali l'ossigeno, la salinità, il pH, il livello di CO_2, possono essere all'origine delle transizioni composizionali trovate tra i pesci (fig. 5).

Sviluppi teorici della teoria neoselezionista

Per riassumere, la teoria neoselezionista: 1) dà una soluzione alla controversia tra neutralisti e selezionisti, poiché concilia la visione neutralista, o piuttosto quasi neutralista, delle mutazioni puntiformi con la selezione al livello regionale; 2) è una teoria con una forte componente epigenomica, in quanto i cambiamenti composizionali del DNA sono accompagnati da cambiamenti nella struttura della cromatina e, di conseguenza, da cambiamenti nell'espressione dei geni, il che conduce a una diminuzione della fitness dei portatori e a una selezione negativa della loro progenie; è importante sottolineare che invocare modificazioni regionali e una partecipazione della cromatina (fig. 2*E*) nell'evoluzione del genoma rappresenta un cambiamento di paradigma rispetto a modelli precedenti che erano centrati sui geni e sulle sequenze regolatrici in senso stretto (promotori); 3) è un'estensione della teoria di Darwin; in realtà, la teoria neoselezionista può essere considerata come una teoria ultradarwiniana, poiché su tempi lunghi anche le mutazioni neutre o quasi neutre sono in definitiva controllate regionalmente in termini composizionali dalla selezione naturale. La teoria neoselezionista ci riporta di conseguenza dalla sopravvivenza del più fortunato alla sopravvivenza del più adatto.

Come ogni buona teoria, anche la teoria neoselezionista permette alcune predizioni. La prima è che differenze nel fenotipo del genoma potrebbero ritrovarsi anche in diverse popolazioni umane. Questa predizione è stata confermata dalla nostra dimostrazione che non solo esistono differenze tra uno specifico genoma umano (il genoma di Craig Venter) e il genoma umano di riferimento, ma che queste differenze consistono principalmente in inserzioni e delezioni concentrate nel cuore del genoma, caratterizzato da una cromatina aperta. La seconda predizione è che alcune di queste differenze possono alterare la fitness del genoma e causare malattie che non sono genetiche, ma genomiche, in cui l'alterazione può riguardare la struttura della cromatina, senza necessariamente toccare la sequenza dei geni o dei loro promotori.

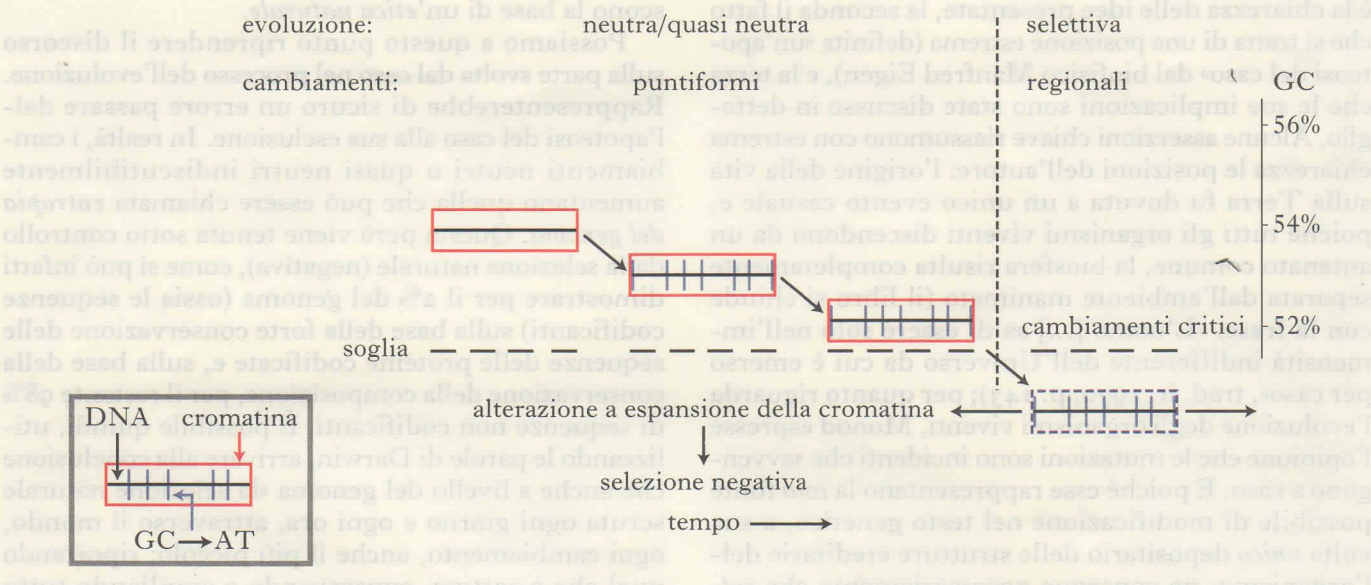

Fig. 6 – Schema della teoria neoselezionista

Decorso nel tempo dei cambiamenti composizionali in una regione ricca in GC del genoma di un vertebrato a sangue caldo secondo il modo conservativo dell'evoluzione (Fonte: figura ridisegnata da Bernardi 2007)

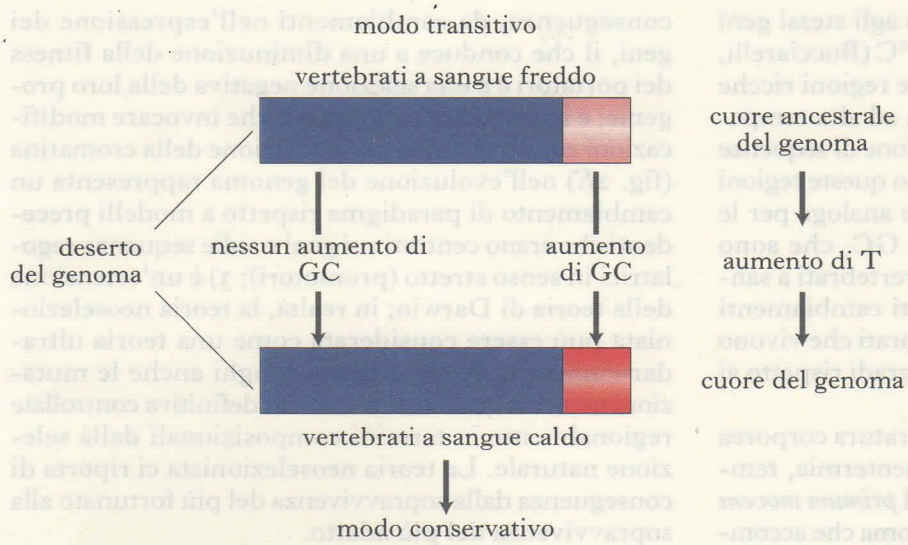

Fig. 7 – Evoluzione composizionale del genoma dei vertebrati

A un modo transitivo, in cui il cuore ancestrale del genoma dei vertebrati a sangue freddo si trasforma nel cuore del genoma dei vertebrati a sangue caldo, succede un modo conservativo, in cui i profili composizionali rimangono costanti (Fonte: figura ridisegnata da Bernardi 2004)

Il caso e la necessità

La prima domanda che dobbiamo porci a proposito della parte del caso nell'evoluzione riguarda l'importanza del problema. Questo interrogativo si può affrontare in diversi modi, ma una via rapida è quella di concentrare l'attenzione sulla posizione presa da J. Monod nel già citato *Le hasard et la nécessité* (1970). Vi sono almeno tre ragioni per questa scelta. La prima è la chiarezza delle idee presentate, la seconda il fatto che si tratta di una posizione estrema (definita «un'apoteosi del caso» dal biofisico Manfred Eigen), e la terza che le sue implicazioni sono state discusse in dettaglio. Alcune asserzioni chiave riassumono con estrema chiarezza le posizioni dell'autore: l'origine della vita sulla Terra fu dovuta a un unico evento casuale e, poiché tutti gli organismi viventi discendono da un antenato comune, la biosfera risulta completamente separata dall'ambiente inanimato (il libro si chiude con la frase: «L'uomo [...] sa di essere solo nell'immensità indifferente dell'Universo da cui è emerso per caso», trad. it. 1970, p. 143); per quanto riguarda l'evoluzione degli organismi viventi, Monod espresse l'opinione che le mutazioni sono incidenti che «avvengono a caso. E poiché esse rappresentano la *sola* fonte possibile di modificazione nel testo genetico, a sua volta *unico* depositario delle strutture ereditarie dell'organismo, ne consegue necessariamente che *soltanto* il caso è all'origine di ogni novità, di ogni creazione nella biosfera» (pp. 95-96). Monod propose inoltre l'idea che «la conoscenza vera ignora i valori» (p. 140) e invocò un'etica della conoscenza, il cui unico valore è rappresentato dalla stessa conoscenza oggettiva.

La visione delle cose a cui noi arriviamo è diversa, se si considerano i seguenti punti: l'origine della vita è vista attualmente come una necessità, tenuto conto delle condizioni fisiche e chimiche prevalenti (Christian René de Duve), piuttosto che come dovuta a un evento eccezionale, a un caso unico; gli organismi viventi sono uniti tra loro da una comune discendenza, punto ovviamente accettato da Monod, e sono modellati dall'ambiente attraverso la selezione naturale (come noi abbiamo dimostrato al livello del genoma). Questa visione propone: 1) l'esistenza di un legame primordiale tra il mondo inanimato dal quale si è originata la vita e il mondo degli organismi viventi; 2) una comune discendenza degli organismi viventi che implica necessariamente un rapporto di fraternità tra di loro; 3) una dipendenza dall'ambiente che comporta nello stesso tempo un rapporto di dipendenza dal mondo inorganico da cui siamo derivati in primo luogo. Mentre il primo punto ci riporta a quella che per Monod era una concezione animistica della natura (che può invece essere considerata come un'antichissima intuizione dei legami, ora confermati in ambito scientifico, tra organismi viventi e natura inanimata), gli altri due punti evidentemente rappresentano dei *valori*, cioè costituiscono la base di un'*etica naturale*.

Possiamo a questo punto riprendere il discorso sulla parte svolta dal caso nel processo dell'evoluzione. Rappresenterebbe di sicuro un errore passare dall'apoteosi del caso alla sua esclusione. In realtà, i cambiamenti neutri o quasi neutri indiscutibilmente aumentano quella che può essere chiamata *entropia del genoma*. Questa però viene tenuta sotto controllo dalla selezione naturale (negativa), come si può infatti dimostrare per il 2% del genoma (ossia le sequenze codificanti) sulla base della forte conservazione delle sequenze delle proteine codificate e, sulla base della conservazione della composizione, per il restante 98% di sequenze non codificanti. È possibile quindi, utilizzando le parole di Darwin, arrivare alla conclusione che anche a livello del genoma «la selezione naturale scruta ogni giorno e ogni ora, attraverso il mondo, ogni cambiamento, anche il più piccolo; riportando quel che è cattivo, conservando e ampliando tutto quello che è buono; lavorando silenziosamente e insensibilmente, quando e dove ce n'è l'opportunità, al miglioramento di ogni essere organico in relazione alle sue condizioni di vita organiche e inorganiche».

Bibliografia

T. Ohta, *Near-neutrality in evolution of genes and gene regulation*, «PNAS. Proceedings of the National academy of sciences of the USA», 2002, 25, pp. 16.134-37.

S. Saccone, C. Federico, G. Bernardi, *Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds*, «Gene», 2002, 1-2, pp. 169-78.

G. Bernardi, *Structural and evolutionary genomics. Natural selection in genome evolution*, Amsterdam 2004.

M. Costantini, O. Clay, F. Auletta, G. Bernardi, *An isochore map of human chromosomes*, «Genome research», 2006, 4, pp. 536-41.

G. Bernardi, *The neoselectionist theory of genome evolution*, «PNAS», 2007, 20, pp. 8385-90.

M. Costantini, F. Auletta, G. Bernardi, *Isochore patterns and gene distributions in fish genomes*, «Genomics», 2007, 3, pp. 364-71.

M. Costantini, O. Clay, C. Federico et al., *Human chromosomal bands*, «Chromosoma», 2007, 1, pp. 29–40.

M. Costantini, M. Di Filippo, F. Auletta, G. Bernardi, *Isochore pattern and gene distribution in the chicken genome*, «Gene», 2007, 1-2, pp. 9-15.

G. Bucciarelli, M. Di Filippo, D. Costagliola et al., *Environmental genomics: a tale of two fishes*, «MBE», 2009, 6, pp. 1235-43.