

Albero filogenetico in forma circolare di 194 specie viventi i cui genomi sono stati interamente sequenziati. L'altezza delle barre all'esterno della circonferenza è proporzionale al numero di geni annotati nel genoma corrispondente. La maggior parte dei genomi sequenziati appartiene a Batteri. La porzione superiore sinistra dell'albero comprende Archaea e Eucarioti, fra cui il riso (*Oryza sativa*, pallino grigio) e l'uomo (*Homo sapiens*, pallino nero).

L'EVOLUZIONE MOLECOLARE

uno sguardo nuovo sui viventi e sulla loro storia

di Giorgio Dieci* e Paolo Tortora**

Che rapporto c'è tra le conoscenze sui viventi acquisite nell'era della post-genomica e le teorie dell'evoluzione? Che cosa si è scoperto a livello molecolare sui meccanismi che hanno dato origine alla varietà degli organismi che oggi popolano la Terra? Soprattutto nei testi scolastici, la questione viene risolta affermando semplicemente che l'evoluzione molecolare ricalca e conferma quella ipotizzata da Darwin per le specie. Alla luce dei più recenti sviluppi della biologia molecolare, questo contributo ripropone il problema in tutta la sua reale profondità. Presentando punti nodali acquisiti o ancora irrisolti degli aspetti molecolari dei viventi e della loro storia evolutiva, esso invita a un livello di comprensione a cui si può tendere solo accettando la sfida della complessità che caratterizza la vita.

La storia della vita sulla Terra appare a noi, quando tentiamo di ricostruirla, come un lentissimo ma inarrestabile dipanarsi di forme materiali diversissime fra loro, ma tutte fondate su di una comune logica molecolare. Per questo, l'indagine scientifica incentrata sulla tessitura molecolare dei sistemi viventi ha dato, negli ultimi cinquant'anni, contributi enormi alla comprensione della dinamica dei processi evolutivi. Lo scenario che sta prendendo forma, soprattutto a seguito dei recenti sviluppi della biologia post-genomica, offre risposte soddisfacenti a molti interrogativi fondamentali ma, nello stesso tempo, sconcerta per la profondità delle nuove prospettive che schiude al nostro sguardo.

Che cosa ci raccontano le macromolecole biologiche

Ci sono, come è evidente, differenze enormi fra un essere umano e un moscerino, fra un topo e il lievito di birra, fra una gallina e una pianta di tabacco. Ma - ed è uno dei fatti più grandiosi e sorprendenti della biologia - gli apparati macromolecolari che plasmano e costituiscono la tessitura intima di ogni cellula di questi esseri viventi, e che garantiscono la loro perpetuazione attraverso i millenni, presentano fra di loro un impressionante grado di somiglianza.

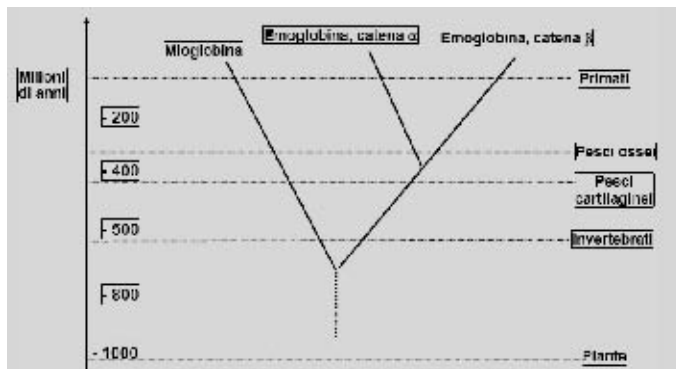
.....
*Professore Associato di Biochimica all'Università degli Studi di Parma.

**Ordinario di Biochimica all'Università degli Studi di Milano-Bicocca.
.....

Nel contempo, però, vengono rivelate in molti casi divergenze anche profonde nell'assetto molecolare dei diversi viventi. Nei mammiferi e nell'uomo, per esempio, ci sono molte proteine che non esistono nel lievito, e molte altre per cui la somiglianza appare vaga, oppure limitata a tratti molto brevi delle proteine stesse. L'osservazione, ripetuta ormai per migliaia di proteine in centinaia di organismi diversi, di gradi diversi di somiglianza nell'ambito di una riconosciuta omologia - somiglianza spesso tanto maggiore quanto più affini morfologicamente appaiono gli organismi - trova la più elegante spiegazione in un quadro di evoluzione macromolecolare che avrebbe accompagnato, e in larga misura anche generato, i cambiamenti evolutivi riscontrabili a livello macroscopico negli esseri viventi del presente e del passato. L'immagine a lato intende chiarire questo concetto mostrando l'albero evolutivo delle globine, così come è stato ricostruito in base al confronto delle loro sequenze.

Tali proteine hanno un'origine molto antica, e la loro funzione (il legame e il trasporto dell'ossigeno all'interno dell'organismo) è essenziale negli organismi pluricellulari di una certa complessità. In particolare la mioglobina immagazzina ossigeno nelle cellule muscolari, l'emoglobina, composta di quattro catene uguali due a due (due α e due β), trasporta ossigeno nel sangue. In base al confronto delle sequenze amminoacidiche e a evidenze paleontologiche indipendenti è stata formulata l'ipotesi dell'orologio molecolare (*molecular clock*), proposta da Linus Pauling e altri scienziati negli anni Sessanta del XX secolo. Essa si può enunciare nel modo seguente: «per una data proteina, il tasso di evoluzione molecolare (= mutazioni accumulate per unità di tempo) è approssimativamente costante in tutte le linee evolutive». Ciò ha portato a ricostruire alberi evolutivi come quello mostrato sopra. Nel caso delle globine il tasso evolutivo è pari all'1% dei residui ogni 5.8 milioni di anni; dunque, se due sequenze divergono del 60%, ciò significa, secondo l'ipotesi, che la divergenza data a $60 \times 5.8 = 348$ milioni di anni fa. Questi stessi studi hanno tuttavia indicato che tra proteine diverse il tasso evolutivo può differire in modo drastico (una variazione di sequenza dell'1% può infatti richiedere un tempo compreso tra 1 e 20 milioni di anni, a seconda dei casi). Ciò sembra dovuto al fatto che proteine diverse possiedono una differente frazione di amminoacidi che tollera mutazioni senza la compromissione della funzione.

Sebbene questa teoria sia stata sottoposta a qualche critica negli aspetti di dettaglio, nelle sue linee essenziali è oggi generalmente accettata, e grazie a essa è possibile stabilire la data approssimativa in cui ha avuto



L'albero evolutivo delle globine
Sulla scala verticale sono indicati i tempi dell'evoluzione.

Corrispondentemente viene indicata l'epoca di comparsa di importanti raggruppamenti biologici (piante, invertebrati eccetera).

La scala orizzontale indica la distanza evolutiva tra diverse proteine nelle diverse epoche (quantificabile in base alla percentuale di residui mutati nelle sequenze). Per esempio, la divergenza tra mioglobina e emoglobina data a circa 500 milioni di anni fa, approssimativamente in corrispondenza della comparsa degli invertebrati. La divergenza tra catene α e β della emoglobina data a poco meno di 400 milioni di anni fa, poco prima della comparsa dei pesci ossei.

luogo la divergenza evolutiva delle diverse specie e, di conseguenza, delle proteine che ciascuna di esse possiede. Ma l'aspetto che maggiormente colpisce nel confronto di proteine correlate nella sequenza, come è il caso delle globine, è il fatto che *la loro struttura tridimensionale è virtualmente identica*, indipendentemente dalla distanza cronologica dell'evento evolutivo che secondo l'ipotesi ha portato alla loro separazione. Ciò rappresenta l'evidenza più schiacciante del fatto che tutte queste proteine derivano da un'unica molecola progenitrice a seguito di un processo evolutivo che le ha differenziate in diversa misura le une dalle altre.

Questo esempio documenta con grande evidenza che è proprio al livello molecolare che si possono comprendere al meglio i meccanismi che stanno alla base del processo evolutivo. Ciò sarà illustrato in maggior dettaglio nelle sezioni successive.

Le basi molecolari dell'ereditarietà e dell'evoluzione

Non c'è evoluzione senza ereditarietà - cioè senza possibilità di trasferimento fedele di informazioni da una generazione alla successiva - e senza la comparsa continua di variazioni nelle informazioni trasferite. Nell'universo biologico, il sistema fondamentale di trasferimento transgenerazionale delle informazioni è quello genetico (anche se, come discusso in una sezione successiva, si possono individuare importanti dimensioni non genetiche dell'ereditarietà). A coronamento dei molti sforzi tesi a svelare il meccanismo del trasferimento genetico di informazioni, attorno alla metà del XX secolo fu chiaro dove questo andasse innanzitutto cercato: nel DNA e nella sua duplicazione, cioè nella possibilità che esso ha di dirigere la sintesi di nuove molecole identiche a sé [1, 2].

Il DNA come molecola dell'informazione genetica ereditabile

Il DNA, come rivelato dal suo utilizzo nei sistemi viventi, è realmente una molecola straordinaria: da un lato esso può contenere, ultimamente scritti nella sua sequenza nucleotidica, i diversi livelli di informazione che sono fondamentali per la formazione, l'auto-mantenimento e il funzionamento di cellule e organismi; dall'altro, come ricordato, esso può propagarsi identico da una generazione all'altra a livello sia cellulare sia di organismo.

L'informazione contenuta nel DNA si attualizza in funzioni precise mediante i processi biochimici di *trascrizione* e di *traduzione*. La trascrizione genera molecole di RNA la cui sequenza nucleotidica è complementare a quella di un tratto di DNA che funge da stampo. Esistono molti tipi diversi di molecole di RNA. La classe più numerosa è costituita dagli mRNA, la cui funzione è strettamente connessa al processo di traduzione. In questo processo, una molecola di mRNA viene letta come serie di triplette

nucleotidiche, a ognuna delle quali corrisponde, secondo il codice genetico universale, uno e uno solo dei venti amminoacidi. Il prodotto della traduzione è quindi una proteina la cui sequenza amminoacidica corrisponde alla sequenza di triplette nucleotidiche dell'mRNA e, di rimando, alla sequenza del DNA originariamente trascritta in mRNA. Ciò che avviene, in definitiva, è che una particolare «lettura» chimica di una sequenza di DNA «esprime» una proteina, cioè un polimero di tutt'altra natura: questa è l'acrobazia chimica, innumerevoli volte ripetuta, su cui poggia l'esistenza di ogni vivente. Le proteine, infatti, sono a fondamento di ogni struttura e attività biologica.

L'insieme dei passaggi che conducono dal DNA alla proteina viene spesso chiamato *espressione genica*, intendendosi comunemente per *gene* la porzione di DNA che può esprimere una proteina. E il gene, naturalmente, è anche un pacchetto discreto, un «quanto» di informazione genetica che può essere trasferito da una generazione all'altra. E, fatto tanto raro quanto importante, ogni gene può subire alterazioni della sua sequenza nucleotidica - *mutazioni* - che possono essere trasferite alle generazioni successive, le quali possono così acquisire la capacità di esprimere proteine ad alterata sequenza e funzionalità (vedi anche la sezione successiva). Considerando che nelle cellule di ogni vivente possono essere espresse migliaia di proteine diverse (circa 6000 nel lievito di birra, almeno 30000 nell'uomo), e che tutti i corrispondenti geni possono andare incontro a mutazione, si riesce facilmente a immaginare uno scenario caratterizzato da un numero pressoché infinito di diversi patrimoni genici, e quindi proteici. All'interno di tale scenario, la selezione naturale imporrebbe una maggiore propagazione transgenerazionale di alcuni arsenali genici/proteici rispetto ad altri, e questo sarebbe l'innescò originario di ogni cambiamento evolutivo. Il cambiamento delle frequenze geniche nell'ambito delle singole popolazioni, mediato dalla selezione generazione dopo generazione, è stato e viene continuamente messo in evidenza dalla *genetica di popolazioni*. Tuttavia, un problema molto dibattuto è se l'intera storia evolutiva dei viventi, come la conosciamo grazie soprattutto alla dimensione di profondità temporale fornita dalla paleontologia, possa essere dedotta in modo soddisfacente dall'azione della selezione naturale sulla variazione genetica disponibile [3]. La selezione naturale, infatti, non si attua su genomi, ma su organismi; il substrato della selezione - e quindi dell'evoluzione - sono i fenotipi, piuttosto che i genotipi. Ed è noto che l'evoluzione fenotipica può procedere in modo fortemente disgiunto dal modo in cui procede la variazione delle sequenze genomiche. Gli anfibi e i mammiferi, per esempio, derivano da un progenitore acquatico comune. Sebbene la velocità di accumulazione delle mutazioni sia la stessa nei genomi di anfibi e mammiferi, i primi si sono evoluti molto più lentamente dei secondi. In altri termini, l'evoluzione molecolare è stata simile nei due gruppi, mentre l'evoluzione fenotipica è stata molto diversa [4].

Le basi molecolari della variazione genetica

Quali sono gli eventi molecolari in grado di generare, a livello dell'informazione contenuta nel DNA, quella varietà necessaria perché si possa avere una selezione naturale?

Agli inizi del XX secolo, lo studio della gametogenesi e della meiosi, e in particolare del processo di ricombinazione del materiale genetico (*crossing over*) che avviene durante la meiosi, suggerirono come sempre nuove variazioni del patrimonio genico potessero essere realizzate nell'ambito di una popolazione. Fu presto chiaro, tuttavia, che in assenza di altri fattori tale ricombinazione non sarebbe in grado di cambiare significativamente il *pool* genetico di una popolazione. La fonte primaria e indispensabile di ogni novità e variazione genetica è un'altra: la mutazione, la quale può essere fissata nell'avventura temporale di una specie se avviene nelle cellule della linea germinale e non produce svantaggi all'individuo che la porta.



Barbara McClintock (1902-1992).

Una delle più grandi figure della biologia del XX secolo. Studiando il mais, scoprì negli anni Quaranta il fenomeno della trasposizione genetica, in seguito riconosciuto fondamentale per l'evoluzione dei genomi. Per questo le fu assegnato il premio Nobel per la medicina e la fisiologia nel 1983.

La possibilità di mutazione è intrinseca alle caratteristiche biochimiche del DNA. Le mutazioni possono innanzitutto presentarsi come conseguenze di alterazioni chimiche del DNA, che ne modificano la sequenza. Tali alterazioni possono essere spontanee, o causate da agenti chimico-fisici esterni, oppure anche da molecole generate nel corso di processi fisiologici. Ma mutazioni possono insorgere anche come errori durante i processi biochimici che interessano il DNA: errori di copiatura durante la replicazione del DNA; mancata separazione di cromosomi che può portare a duplicazioni genomiche totali o parziali; varie tipologie di eventi ricombinativi, quali la duplicazione di segmenti di DNA e la trasposizione, che sono noti aumentare la probabilità di generazione di nuovi geni.

Malgrado la diversità dei meccanismi ipotizzati per questi processi, il loro risultato è sempre quello di aumentare la complessità genica degli organismi. In particolare, mediante eventi di «*crossing over* ineguale», si possono originare nuovi geni identici a geni pre-esistenti, i quali poi possono trasformarsi, mediante mutazioni successive, in nuovi geni a funzione distinta. È così che si sono generate, per esempio, le famiglie dei geni per le globine (si veda pagina 9), o per le proteine *homeobox* che regolano i processi di sviluppo e differenziamento di moltissimi organismi pluricellulari. Mediante simili eventi ricombinativi possono essere generate anche combinazioni sempre nuove di un numero relativamente piccolo di diversi segmenti di DNA corrispondenti a precisi domini funzionali di proteine. Questi processi di riassortimento di domini, globalmente definiti *domain shuffling* o *exon shuffling*, sono all'origine della grandissima varietà di proteine modulari di cui sono ricchi gli eucarioti superiori [5, 6].

DOMINI PROTEICI

All'osservazione della loro struttura tridimensionale, molte proteine appaiono come un'unica entità globulare. Altre proteine appaiono composte da diverse entità globulari unite l'una all'altra da tratti estesi e ad andamento irregolare. Ciascuna entità globulare è detta dominio, una traduzione approssimativa del termine inglese domain. Nelle proteine composte di molteplici domini ciascuno di essi si fa carico di un compito specifico nell'ambito della funzione complessiva. Dunque la struttura a dominio si può definire modulare, dove i singoli moduli sono appunto i domini. A titolo esemplificativo la piruvato chinasi è un enzima glicolitico che catalizza la formazione del piruvato: esso consiste di un dominio regolatorio, un dominio che lega una delle due molecole partecipanti alla reazione (il fosfoenolpiruvato) e un dominio che lega l'altro reagente (l'ADP).

La struttura a dominio è diffusa in tutto il mondo biologico, ma è soprattutto frequente negli organismi eucariotici, dove alcune proteine consistono addirittura di parecchie decine di domini, sebbene normalmente il loro numero si limiti a poche unità.

Anche la genesi delle proteine multidominio può essere adeguatamente compresa in una prospettiva evolutiva. I singoli domini sono probabilmente esistiti in origine come proteine singole. Grazie a eventi di riassortimento casuale del DNA, si sarebbero creati dei geni codificanti proteine formate da molteplici domini. Le proteine generate in tal modo, se dotate di una funzione vantaggiosa, avrebbero avuto successo evolutivo trasmettendosi così alla progenie. Ciò spiegherebbe perché un determinato dominio si possa trovare in proteine differenti e perché al contempo nelle diverse proteine contenenti un dato dominio la struttura complessiva di questo sia sostanzialmente la stessa, e analogamente la sua sequenza amminoacidica sia simile anche se non identica. Nel corso dell'evoluzione si realizza infatti una divergenza delle sequenze a causa dell'accumulo progressivo di mutazioni casuali.

La storia evolutiva che ha portato alla formazione di una proteina composta di più domini trova riscontro anche nell'organizzazione del genoma eucariotico, dove accade di osservare, almeno in certi casi, che un singolo dominio sia codificato da un esone, e che quindi tratti codificanti domini adiacenti siano intervallati da introni. In breve, la corrispondenza di un dominio ad un determinato esone è essa stessa rappresentativa dell'autonomia evolutiva dei domini, in quanto essi si presentano come entità distinte anche a livello dell'organizzazione genica. Questo concetto è stato formalizzato con la teoria dell'*exon shuffling* (rimescolamento degli esoni) proposta nel 1977 da Walter Gilbert, secondo la quale esoni dello stesso o di diversi geni codificano ciascuno un dominio, e occasionalmente possono essere rimescolati, producendo così una varietà di diverse proteine.

Pur essendo l'unica fonte di vera novità nei genomi, l'instabilità biochimica del DNA costituisce una continua minaccia per la sopravvivenza degli individui e per il perdurare stesso della vita. Per questo tutti i viventi, dai batteri all'uomo, possiedono sofisticati sistemi enzimatici preposti alla riparazione del DNA danneggiato o alterato. Tali sistemi si basano su dispositivi molecolari per il riconoscimento del danno (per esempio una rottura a singolo o doppio filamento del DNA, oppure l'introduzione in un filamento di un nucleotide non complementare al filamento stampo), a cui fa seguito la messa in atto di processi cellulari che culminano nella riparazione del danno.

Evolvibilità: una questione aperta

È vertiginoso pensare a come la varietà dei viventi e il loro perdurare poggino sulla instabilità del DNA nei sistemi cellulari e, nello stesso tempo, sulla sua stabilità, e come quest'ultima poggia a sua volta su funzioni di riparazione che sono codificate nella sequenza stessa del DNA! La vertigine diviene ancora più grande qualora si consideri la possibilità che la comparsa stessa delle mutazioni sia un processo non totalmente

accidentale, ma in qualche modo regolato dall'organismo stesso. È ben noto che, durante lo sviluppo delle cellule del sistema immunitario, le sequenze di DNA contenute nei geni degli anticorpi subiscono tagli, trasferimenti, giunzioni con altre sequenze al fine di generare una grandissima varietà di possibili sequenze, e quindi di possibili anticorpi.

Questo esempio suggerisce che le mutazioni - materia prima dell'evoluzione - possono verificarsi come risultato non solo di eventi casuali, ma anche di programmi cellulari specifici e controllati, attivati - per esempio - in risposta a particolari stimoli ambientali. La possibilità di una ristrutturazione controllata del patrimonio genetico, cioè di una regolazione della fluidità dei genomi, è direttamente connessa all'idea, affascinante, che anche la capacità di evolvere debba poter essersi evoluta [7].

Ci sembra opportuno sottolineare, a questo punto, come l'insieme di tutte le nostre conoscenze - in molti casi estremamente dettagliate - sulle basi molecolari dell'ereditarietà e della generazione della diversità genetica siano ancora largamente insufficienti a spiegare come sia stata e sia generata quella prodigiosa innovazione morfologica, quel fiorire incessante di forme che è da sempre l'aspetto più provocatorio della storia della vita sulla Terra.

Come è stato innescato, e secondo quali leggi si è snodato innumerevoli volte, il processo di generazione di nuove forme viventi?

Secondo la cosiddetta Sintesi Moderna, che rappresenta il paradigma corrente della moderna biologia evoluzionistica [8], l'evoluzione dei fenotipi è innanzitutto un risultato dell'azione della selezione naturale, mentre la generazione di forme su cui la selezione agisce tende a essere considerata come condizione che è sì necessaria, ma priva di influenza sulla direzione dell'evoluzione.

Tuttavia, l'idea che la selezione naturale sia l'unica forza-guida dell'evoluzione viene attualmente messa in discussione da esponenti di uno specifico e vasto campo di ricerca biologica, che si è rivelato estremamente produttivo negli ultimi vent'anni: la biologia evoluzionistica dello sviluppo (comunemente abbreviata in *evo-devo*).

I ricercatori in ambito *evo-devo* pongono l'accento sul fatto che le caratteristiche specifiche dei processi di sviluppo degli organismi pluricellulari devono aver giocato un ruolo fondamentale nella generazione della diversità di forme viventi. Per usare le parole di un suo autorevole esponente: «*evo-devo* sposta l'attenzione della spiegazione evoluzionistica da ciò che è esterno e contingente a ciò che è interno e inerente; postula che la base causale della forma fenotipica non risieda nella dinamica di popolazione, oppure nell'evoluzione molecolare, ma piuttosto in proprietà che sono intrinseche alle dinamiche di sviluppo degli organismi» [9].

Limiti della visione gene-centrica della vita e dell'evoluzione

Grazie alla sua semplicità concettuale e di applicazione, lo schema «un gene → una proteina → una funzione» è stato ed è tuttora molto accattivante, e quindi largamente impiegato per dipingere e comunicare scenari dell'organizzazione dei viventi e della loro evoluzione.

Ancora oggi, tuttavia, esso è in grado di generare e promuovere visioni improntate a un determinismo genetico/biologico della vita umana che appaiono tanto ottimisticamente ingenuo quanto risuonarono drammaticamente inattendibili in alcune grandi coscienze del XIX secolo.

Si tratta, in realtà, di uno schema riduttivo il quale, se applicato dogmaticamente, finisce col generare scenari fuorvianti, per diverse ragioni che qui riassumiamo, limitandoci agli aspetti molecolari della questione.

È interessante a questo proposito confrontare gli scenari possibili, derivanti dalla conoscenza scientifica dell'uomo, tratteggiati dal famoso biologo cellulare Harvey F. Lodish nel 1995 e da Fëdor Dostoevskij nel 1864. Dichiarò Lodish: «Lo *screening* genetico pre-impianto degli embrioni può divenire presto *routine*. [...] le predizioni riguardo alla struttura e alla funzione delle proteine saranno abbastanza accurate da rendere possibile dedurre, in maniera automatica, le proprietà rilevanti di molti peptidi importanti, così come la regolazione della loro espressione (quanta proteina, ad esempio, verrà prodotta nel corso di una particolare fase dello sviluppo di un determinato tessuto o tipo di cellula) a partire dalla sequenza del solo DNA del genoma. Tutte queste informazioni verranno trasferite a un supercomputer, insieme con notizie di carattere ambientale, tra cui probabilmente l'alimentazione, le tossine ivi presenti, la luce solare e così via. Ciò si tradurrà in un film a colori, che vedrà l'embrione svilupparsi in un feto, venire alla luce e quindi crescere fino a diventare adulto, evidenziando [...] le proprie dimensioni e forme corporee, il colore dei capelli, della pelle e degli occhi. Alla fine la sequenza di DNA sarà estesa a coprire geni rilevanti per caratteri quali la parola e le doti musicali; la madre potrà sentire parlare o cantare l'embrione come se fosse già una persona adulta (Lodish H.F. 1995, *Science*, vol. 267, p. 1609; cit. in E. Jablonka e M. Lamb (2007) *L'evoluzione in quattro dimensioni*, UTET, Torino). Tutt'altro è il tono del personaggio del sottosuolo, più di un secolo prima: «[...] allora, dite voi, la scienza stessa insegnerà all'uomo [...] che in realtà egli non ha né ha mai avuto né volontà né capriccio, e che anche lui non è nulla più che una specie di tasto di pianoforte o di una puntina d'organetto; e che inoltre ci sono al mondo anche le leggi naturali; cosicché, qualunque cosa egli faccia, avviene non già per il suo volere, ma di per sé, secondo le leggi naturali. Per conseguenza, queste leggi naturali basta scoprirle, che l'uomo non risponderà più delle sue azioni, e vivere gli sarà straordinariamente facile. Tutte le azioni umane, va da sé, allora saranno calcolate secondo queste leggi, matematicamente, sul tipo di una tavola di logaritmi fino a 108000, e iscritte nel calendario; o, meglio ancora, compariranno delle pubblicazioni benpensanti, sul tipo degli odierni dizionari enciclopedici, in cui tutto sarà enumerato e segnato in modo così preciso che nel mondo non ci saranno più né azioni né avventure.» (F. Dostoevskij, *Memorie del sottosuolo*, Torino 1942).

Le sorprese dell'RNA

Innanzitutto - e questo è noto da più di mezzo secolo - alcune funzioni cellulari basilari, come la traduzione stessa, coinvolgono svariate molecole di RNA (per esempio rRNA, tRNA, piccoli RNA nucleolari, o snoRNA) che vengono generate mediante trascrizione del DNA, ma non vengono tradotte in proteine, in quanto agiscono come tali.

In altri termini, gli mRNA sono solo una piccola parte del trascrittoma, che include molte altre tipologie di RNA.

Per tenere conto di questi RNA (che vengono definiti ncRNA, per *non-protein-coding* RNA), occorre uno schema modificato:

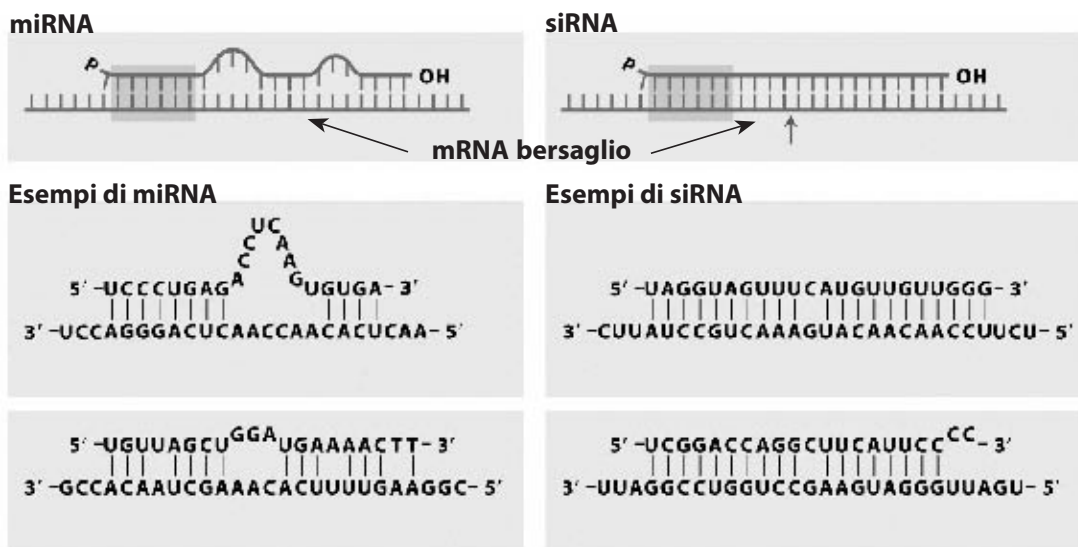
«un gene → una proteina/un ncRNA → una funzione».

La rilevanza di questa modificazione dello schema appare ogni giorno più grande, poiché è sempre maggiore il numero di diversi ncRNA che vengono scoperti svolgere funzioni importanti nelle cellule.

TRASCRIPTOMA

Con questo termine si intende generalmente l'insieme degli RNA che vengono prodotti mediante trascrizione, e successivo processamento post-trascrizionale, a partire da un genoma. Mentre il genoma non cambia da una cellula all'altra dello stesso organismo, e durante le diverse condizioni e fasi di vita di una cellula, il trascrittoma è estremamente dinamico. I profili di trascrizione di un genoma, infatti, cambiano profondamente da un tipo cellulare all'altro, da una condizione ambientale all'altra, da una fase all'altra della vita di una cellula. Esiste, pertanto, un numero elevatissimo di possibili trascrittomi derivanti da un singolo genoma. I trascrittomi possono essere studiati mediante l'utilizzo della tecnologia dei DNA microarrays, che permette di stabilire ad altissima risoluzione i livelli dei trascritti corrispondenti a ogni singola regione del genoma, o, più recentemente, mediante potenti strategie di sequenziamento di nuova generazione (il cosiddetto *deep sequencing*).

Migliaia di specie di ncRNA sono state scoperte negli ultimi anni. Alcuni sono RNA di soli 20-25 nucleotidi (denominati per questo microRNA o miRNA) che operano nella modulazione del processo di traduzione appaiandosi a regioni complementari sulla molecola di mRNA, come mostrato nella figura seguente. Molte altre specie di RNA sono di grandi dimensioni e a funzione sconosciuta.



I miRNA si appaiano in modo imperfetto a una regione parzialmente complementare dell'mRNA bersaglio, causando un'inibizione della traduzione dell'mRNA.

I siRNA (*small interfering RNA*) riconoscono una regione perfettamente complementare dell'RNA bersaglio, e ne causano la degradazione.

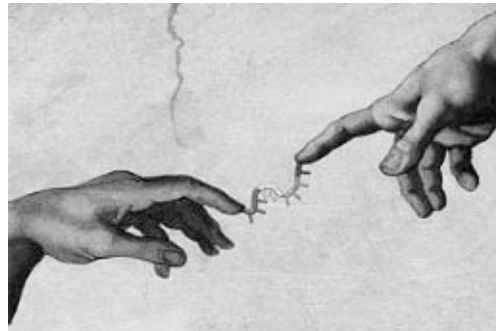
Evidenziata da un rettangolo più scuro è la regione critica per il riconoscimento dell'mRNA da parte del piccolo RNA.

Vastissime regioni di DNA che fino a poco tempo fa si consideravano silenziose dal punto di vista dell'espressione, si sono rivelate sedi di attiva trascrizione, con produzione di innumerevoli nuove specie di RNA alle quali solo ora cominciano a essere attribuite funzioni di grande rilevanza, per esempio nei diversi livelli di organizzazione del genoma.

Nel 2007 sono stati pubblicati dalla rivista *Nature* i risultati della fase pilota dell'ambizioso progetto ENCODE (*encyclopaedia of DNA elements*), il cui fine è identificare nel genoma umano ogni sequenza di DNA con proprietà funzionali. Lo studio iniziale, che ha coinvolto un grande numero di ricercatori di circa 80 laboratori, ha studiato in dettaglio, attraverso una molteplicità di approcci molecolari, l'1% del genoma umano (corrispondente a una sequenza di circa 30 milioni di nucleotidi, o 30 megabasi), generando e analizzando nel complesso circa 600 milioni di dati sperimentali. Ciò che è emerso è l'esistenza di una trascrizione pervasiva che genera un complesso *network* di trascritti fra loro interconnessi, di profonde interazioni funzionali fra processi diversi come la trascrizione e la replicazione del DNA, e di numerosissime sequenze che devono essere catalogate come funzionalmente importanti, anche se la loro funzione è sconosciuta. Sorprendentemente, molti elementi funzionali rivelati da ENCODE sembrano non aver subito costrizioni nel corso dell'evoluzione dei Mammiferi. Questa osservazione suggerisce l'esistenza, nei genomi, di un largo bacino di elementi neutrali che sono biochimicamente attivi, ma non producono specifici vantaggi all'organismo! [The ENCODE Project Consortium (2007) *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature* 447, 799-816]

Basti pensare che i geni che vengono espressi in proteine costituiscono meno del 2% del genoma umano. Eppure, per ragioni ancora sconosciute, la maggior parte del nostro genoma è trascritto in RNA. Ora, l'esistenza di migliaia di prodotti genici che non sono proteine dovrebbe persuaderci a limitare sempre meno a queste ultime i criteri del nostro approccio all'evoluzione molecolare e ai fondamenti molecolari della vita. In particolare, siamo spinti a chiederci: in che misura le mutazioni che alterano gli ncRNA contribuiscono alla generazione della diversità su cui opera la selezione naturale?

Ma gli studi recenti di trascrittomiche, unitamente alle aumentate conoscenze sul processo di *splicing* dell'RNA, hanno reso inaspettatamente difficile la definizione stessa di gene in termini molecolari. È noto da decenni che in quasi tutti gli eucarioti, incluso l'uomo, le proteine hanno origine in molti casi dalla trascrizione di geni discontinui, in cui brevi tratti di sequenza nucleotidica che verranno tradotti in proteina (detti *esoni*), si alternano a tratti anche molto lunghi di sequenza (gli *introni*) che non risulteranno inclusi nell'mRNA e quindi non saranno tradotti. Il processo di rimozione degli introni e ricomposizione degli esoni a generare l'mRNA maturo viene indicato con il termine *splicing*. Ora, in moltissimi casi, per una stessa sequenza trascritta esistono modalità alternative con cui gli esoni possono essere ricomposti, e quindi esistono numerose varianti proteiche generate mediante *splicing* alternativo. È noto l'esempio di una sequenza trascritta del genoma di pollo che possiede 576 varianti alternative di ricomposizio-



L'immagine con cui il settimanale inglese *The Economist* (14 Luglio 2007) ha enfatizzato le nuove scoperte sui piccoli RNA, che hanno cambiato profondamente la nostra percezione dei circuiti regolativi cellulari e del contenuto di informazione dei genomi. L'importanza di tali scoperte è stata riconosciuta anche con il conferimento, nel 2006, del premio Nobel per la medicina e la fisiologia a Craig Mello e Andrew Fire.

ne degli esoni, e la cui espressione può quindi generare altrettante varianti proteiche! Occorre notare, inoltre, che lo *splicing* alternativo riguarda anche la generazione di ncRNA, rendendone ancora più difficile la comprensione. Così, lo schema del passaggio dall'informazione genetica alla funzione dovrebbe diventare:

«un gene \rightarrow n proteine/ n ncRNA \rightarrow n funzioni»,

dove n può essere facilmente un numero dell'ordine delle centinaia. A complicare ulteriormente il quadro, si stanno scoprendo sempre più numerosi esempi della cosiddetta *trascrizione antisense*. Data una sequenza di DNA, tipicamente codificante per n varianti proteiche dovute a *splicing* alternativo, si rilevano molto spesso molecole di RNA espresse mediante trascrizione del filamento opposto del DNA. Gli RNA antisense così generati si possono appaiare agli RNA «senso», o al filamento di DNA complementare, formando degli ibridi RNA:RNA o RNA:DNA che possono avere effetti molteplici sull'espressione stessa del DNA da cui originano. Anche i trascritti antisense, poi, possono subire *splicing* alternativo. Ma non è tutto: alcune molecole di RNA prodotte nelle cellule hanno una sequenza nucleotidica che include tratti corrispondenti a segmenti di sequenza genomica fra di loro lontanissimi, come se il «gene» corrispondente a questi RNA fosse in realtà un'entità composita fisicamente frammentata e dispersa nella vastità del genoma! [10].

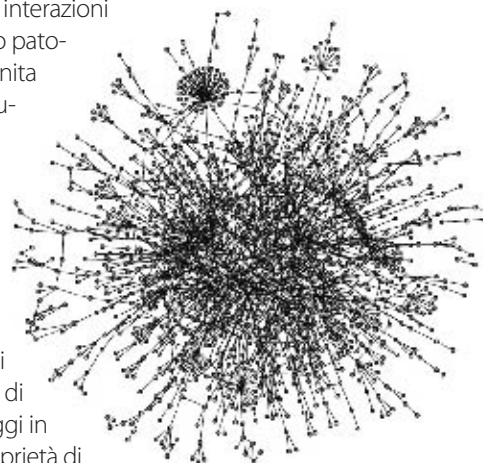
Che quadro emerge dall'osservazione di questi fatti nuovi, di questi scenari che sarebbero parsi inverosimili fino a pochi anni fa? Innanzitutto, è necessario prendere atto che ogni tratto di DNA può essere letto in più di un modo, ed esprimere quindi più di una macromolecola funzionale. Va in questa direzione un'autorevole (e difficoltosa) definizione in termini molecolari di «gene» proposta molto recentemente: «Un gene è una unione di sequenze genomiche codificante per un insieme coerente di prodotti funzionali potenzialmente sovrapposti.» [11]. Come conseguenza, ogni singola mutazione del DNA può avere ripercussioni su molteplici macromolecole tutte derivanti dall'espressione del DNA mutato, e quindi su tutti i processi e le funzioni in cui queste macromolecole sono coinvolte.

PROTEOMA

Questo termine indica l'insieme delle proteine prodotte a partire da un determinato genoma, mediante i processi di trascrizione e successiva traduzione del mRNA. Come il trascrittoma, anche il proteoma è estremamente dinamico, essendo i processi di trascrizione e traduzione profondamente regolati in risposta ai più diversi stimoli. Inoltre, ogni proteina è soggetta a numerose possibili modificazioni post-traduzionali (come fosforilazioni, glicosilazioni, ecc.) che ne modificano la struttura e la funzione, cosicché possono esistere in ogni cellula numerose e diverse forme modificate della stessa proteina. In altri termini, dalla espressione di un gene codificante per una sola proteina, cioè per una sola sequenza amminoacidica, possono in ultima analisi prodursi n molecole proteiche che differiscono fra loro per la presenza o assenza di gruppi chimici sostituenti su particolari amminoacidi. La proteomica, cioè la recente branca della biochimica che si occupa dello studio su larga scala dei proteomi, si basa in larga misura sui moderni sviluppi della spettrometria di massa in combinazione con la disponibilità sempre crescente di sequenze genomiche complete.

Entra in scena l'interattoma

Ma lo schematismo riduttivo «un gene → una proteina → una funzione» vacilla non solo a causa delle nostre aumentate conoscenze sul continente sommerso del ncRNA. Negli ultimi dieci anni è emerso con evidenza sempre maggiore che le funzioni biologiche di un organismo non sono la semplice sommatoria delle funzioni di cui le singole molecole - in particolare le proteine - si fanno carico. Al contrario, sta diventando sempre più chiaro che le proteine realizzano le loro funzioni per lo più cooperando tra di loro mediante una rete estremamente complessa di interazioni, ciascuna delle quali attuata mediante il legame fisico di una proteina all'altra. Ciascuno di questi eventi di legame porta all'attivazione o all'inibizione di specifiche funzioni, cosicché l'*output* complessivo è dato dalla modalità con cui l'insieme di tali interazioni viene modulato in dipendenza da uno specifico stato fisiologico o patologico. In altre parole, a seconda delle condizioni, una parte definita di questo complesso circuito viene acceso o spento, e ciò si traduce in un determinato stato funzionale della cellula o dell'organismo. L'insieme delle interazioni che le proteine di un organismo possono instaurare tra di loro è detto interattoma. Un esempio nell'immagine a lato. In termini più generali, lo studio dell'interattoma è una branca della biologia dei sistemi (si veda riquadro alla pagina successiva). Con tale espressione si intende un approccio interdisciplinare che studia sistematicamente le interazioni complesse dei sistemi biologici e che quindi coinvolge tutti i componenti molecolari e non soltanto le proteine. Ogni tentativo di interpretazione riduzionistica degli organismi si imbatte quindi oggi in formidabili difficoltà, quando avanzi la pretesa di giustificare le proprietà di questi semplicemente in base alla conoscenza delle proprietà delle parti che li costituiscono [12]. Coerentemente, anche il nostro modo di guardare all'evoluzione dovrebbe sempre più tener conto della peculiare complessità degli organismi viventi, la quale è resa ancora più evidente dalla nostra crescente consapevolezza delle componenti non genetiche dell'evoluzione.



L'interattoma di lievito

Ogni circolo rappresenta una proteina e i trattini le interazioni fra di esse.

Tratto da Dunker, A.K., Cortese, A.S., Romero, P., Iakoucheva, L.M., Uversky, V.N. (2005) FEBS J., 272, 5129-5148.

La dimensione epigenetica dell'ereditarietà e dell'evoluzione

L'esistenza e la grande rilevanza delle dimensioni non genetiche dell'ereditarietà possono essere apprezzate grazie a un semplice esempio. Il nostro fegato è costituito da milioni di cellule epatiche le quali, essendo derivate da cellule progenitrici comuni, sono tutte estremamente simili fra di loro per morfologia e caratteristiche metaboliche. Esse però sono molto diverse dai cardiomiociti che costituiscono il muscolo cardiaco, e dai più svariati e specializzati tipi cellulari che ritroviamo nel rene, nel sangue, nel sistema nervoso, eccetera, i quali a loro volta tendono a presentarsi come linee di cellule identiche. Il fatto più sorprendente, qualora si rifletta sulla diversità dei tipi cellu-

BIOLOGIA DEI SISTEMI

La biologia dei sistemi si prefigge di comprendere e predire le proprietà degli organismi in base alla conoscenza dell'insieme dei componenti molecolari di cui consistono e delle interazioni dinamiche, cioè variabili nel tempo, che questi attuano tra di loro. Non si occupa quindi dei singoli meccanismi molecolari: il suo approccio è olistico e non riduzionistico. In tal senso si tratta di una visione sostanzialmente nuova dei meccanismi di funzionamento dei sistemi biologici. La biologia dei sistemi ha avuto un grande impulso in tempi recenti (soprattutto gli ultimi dieci anni) principalmente grazie a sviluppi tecnologici che hanno reso praticabili sperimentalmente principi che in passato avevano una valenza essenzialmente speculativa. Le discipline recenti che rappresentano il presupposto tecnologico della biologia dei sistemi sono specialmente (ma non esclusivamente) la genomica, la trascrittomica, la proteomica e l'interattomica. La genomica e la trascrittomica si sono sviluppate grazie all'impulso ricevuto dall'enorme progresso delle tecniche di sequenziamento del DNA e di produzione di *microarray*, grazie alle quali è stato possibile identificare e sequenziare l'insieme di tutti i geni di un organismo (genoma) e l'insieme dei geni trascritti in una data condizione fisiologica (trascrittoma). La proteomica e la interattomica si sono sviluppate grazie a perfezionamenti delle tecniche della spettrometria di massa, che oggi consente di identificare idealmente tutte le proteine presenti in un organismo nei diversi stati funzionali (proteoma) e le interazioni che esse instaurano tra di loro e con gli acidi nucleici (interattoma). Inoltre, la predizione del comportamento di un organismo sulla base della conoscenza di tutte le sue componenti molecolari (in particolare proteine e acidi nucleici) e della loro rete di interazioni, richiede modelli predittivi che vengono sviluppati mediante approcci della teoria dei sistemi, della bioinformatica e della matematica-statistica. Sebbene a oggi i risultati conseguiti dalla biologia dei sistemi siano parziali, è ragionevole prevedere significativi progressi negli anni a venire, anche se verosimilmente la comprensione dei sistemi biologici resterà sempre incompleta. Le ricadute più significative potrebbero portare alla predizione degli effetti complessivi dei trattamenti farmacologici e delle patologie di cui si conoscano i fondamenti molecolari.

lari presenti negli organismi complessi, è che il patrimonio genetico di ogni cellula è - salvo rarissime eccezioni note - esattamente lo stesso. La differenza fra un epatocita e un neurone non è dovuta a un diverso patrimonio genetico, ma piuttosto a un diverso profilo di espressione dello stesso genoma. Si tratta, cioè, di una differenza non genetica, ma epigenetica. Alcuni geni sono espressi più nell'epatocita che nel neurone, o viceversa, altri sono molto espressi in entrambi, altri ancora in nessuno dei due tipi cellulari. Ognuno delle molte migliaia di geni contenuti nel genoma (siano essi codificanti per proteine o per RNA non tradotti) avrà un certo livello di espressione, diverso in ogni tipo cellulare e in ognuna delle condizioni ambientali e fasi della vita in cui le cellule possono trovarsi. Inoltre - ed è il fatto più rilevante per la nostra discussione - quando le cellule epatiche si dividono danno origine ad altre cellule epatiche, e lo stesso accade per le cellule dell'epidermide, del rene, eccetera. Nel corso del loro sviluppo, queste cellule hanno acquisito informazioni non genetiche che possono trasmettere alla progenie. Quello descritto è l'esempio più immediato e imponente di sistema ereditario epigenetico. Ora, una grande domanda che comincia a essere posta in modo sempre più autorevole è: in che misura i sistemi ereditari epigenetici (e, più in generale, i sistemi ereditari non genetici) contribuiscono all'evoluzione? Si tratta di una domanda di grandissima rilevanza, poiché variazioni di stati epigenetici, e i conseguenti fenotipi, possono manifestarsi negli individui in risposta a stimoli ambientali, e quindi una loro trasmissione alla progenie si configurerebbe come ereditarietà di caratteri acquisiti - un'idea che appare ancora largamente eretica rispetto all'attuale pensiero evolutivista. Eva Jablonka e Marion Lamb hanno recentemente fornito una esauriente e ben documentata tratta-

zione (a cui senz'altro si rimanda) dei sistemi ereditari epigenetici e del loro potenziale impatto sull'evoluzione [13]. Fra questi, possono essere distinti: i) l'eredità di modelli di attività genica, che trovano una limpida esemplificazione nella biologia molecolare del batteriofago *Lambda* e del suo ciclo vitale [14]; ii) l'eredità di particolari sistemi di marcatura della cromatina, che consentono il perpetuarsi di specifici stati di attività o inattività dei geni; iii) il silenziamento a RNA (o RNAi, abbreviazione di *RNA interference*), che può produrre una inibizione ereditabile dell'espressione di geni specifici; iv) l'eredità strutturale, in base alla quale una struttura cellulare con una particolare configurazione (per es. un tratto di membrana plasmatica con certe caratteristiche) può guidare la formazione di una struttura identica a sé nella cellula figlia. Uno degli esempi più impressionanti di eredità strutturale è costituito dalle proteine prioniche, una delle quali è alla base del morbo della mucca pazza (BSE, *Bovine Spongiform Encephalopathy*). Queste proteine possono esistere in conformazioni alternative, e molecole proteiche in una delle due conformazioni possono indurre molecole in conformazione alternativa ad assumere la propria. La scoperta di questa forma di eredità proteica per contiguità è valsa a Stanley Prusiner il premio Nobel per la Medicina e la Fisiologia nel 1997. La possibilità di una eredità proteica e, più in generale, di una eredità strutturale dovrebbe spingerci a non dimenticare un principio fondamentale, mai smentito, della Biologia: il fatto che ogni cellula può derivare solo da altre cellule, la sua membrana da una membrana pre-esistente, i suoi mitocondri da altri mitocondri, e questi a loro volta da altri precedenti, e le cellule più antiche da cellule precedenti, e così via fino a risalire ai tempi remoti e oscuri in cui l'avventura è cominciata. In termini pratici, non si è mai vista una cellula auto-costruirsi a partire dal suo DNA. L'attenzione a questo fatto conduce a una considerazione semplice, ma forse di grande importanza per il nostro sguardo sull'eredità e sull'evoluzione: il vero stampo su cui si modella, si costruisce ogni cellula è, nella sua interezza, la cellula-madre da cui essa origina, così come ogni vivente prende forma a partire da cellule germinali di viventi simili a sé. Il carattere pervasivo e l'inestricabilità, la costanza e l'onnipresenza di ciò che, nelle cellule, non è geneticamente codificato, ne fanno qualcosa di sfuggente, di inaccessibile all'indagine analitica: qualcosa che potrebbe svolgere ruoli fondamentali, ma a nostra insaputa; qualcosa di perfettamente nascosto perché è dappertutto, sempre.

Conclusioni

In conclusione, ci sembra importante sottolineare come le nostre conoscenze sui possibili meccanismi molecolari dell'evoluzione si siano negli ultimi anni enormemente ampliate, e come questo allargarsi dell'orizzonte conoscitivo abbia portato con sé, come prima e più feconda conseguenza, nuove impellenti domande su aspetti fondamentali della vita e della sua storia sulla Terra. Un primo ordine di domande, impostosi proprio in concomitanza con l'inizio dell'era post-genomica della biologia, deriva dalla constatazione che la nostra «lettura»

umana delle informazioni scritte nel DNA di un genoma ci rivela solo una piccola parte di ciò che la lettura reale, chimica di esso è in grado di esprimere da miliardi di anni. Che cosa accade nelle innumerevoli cellule di un organismo complesso, in funzione delle molteplici diverse letture chimiche del genoma, letture a loro volta condizionate, secondo dinamiche ancora incomprese, da un larghissimo spettro di variabili non-genetiche? Questo è il tipo di domande attraverso le quali il biologo post-genomico è ormai costretto a interrogarsi, quando volge il proprio sguardo ai genomi senza astrarli dal loro contesto vivente. Un secondo ordine di domande, caratteristico della nuova biologia dei sistemi, nasce dalla constatazione che le proprietà fondamentali e distintive delle cellule e degli organismi emergono da *network* di interazioni fra insiemi di proteine, e fra proteine e altre molecole, secondo forme e criteri che sfuggono agli approcci riduzionistici. L'irriducibilità delle proprietà emergenti dei sistemi biologici rende evidente, fra l'altro, che l'evoluzione di tali sistemi non può essere indagata in modo esauriente in termini di frequenze alleliche considerate individualmente, al di fuori dei *network* di interazioni fra geni. Un terzo ordine di domande, che sempre più insistentemente chiedono di essere affrontate, riguarda i sistemi di eredità (e quindi di possibile evoluzione) non genetica, che includono non soltanto le modalità epigenetiche di trasmissione dei caratteri, ma anche quelle comportamentali e simboliche, di cui l'evoluzione culturale dell'uomo è profondamente intrisa. Infine, non si può non rimanere stupiti da come la varietà e la bellezza delle manifestazioni della vita sulla Terra abbiano avuto origine da un intrecciarsi di eventi il cui connotato principale è una sostanziale imprevedibilità. L'evoluzione non è certamente un processo deducibile in modo esatto da una conoscenza analitica delle premesse, per quanto dettagliata e precisa possa essere tale conoscenza. La vita e la sua storia si presentano a noi, anche dopo tre secoli di sviluppo rigoglioso delle scienze biologiche, con le sembianze di un fatto straordinario, irripetibile, a cui abbiamo il privilegio di partecipare con la gratitudine e lo stupore di invitati dell'ultima ora. ❖

Riferimenti per l'immagine a pagina 6 - http://en.wikipedia.org/wiki/File:Tree_of_life_with_genome_size.svg.
La fonte di Wikipedia è: <http://itol.embl.de/itol.cgi> Basato sull'articolo: Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) *Toward automatic reconstruction of a highly resolved tree of life*. *Science* 311:1283-1287.

INDICAZIONI BIBLIOGRAFICHE

- [1] Watson J.D., Crick F.H.C., *A structure for deoxyribose nucleic acid*, *Nature* 171 (1953) 737-738.
- [2] Watson J.D., Crick F.H.C., *Genetical implications of the structure of deoxyribonucleic acid*, *Nature* 171 (1953) 964-967.
- [3] Eldredge N., *Ripensare Darwin. Il dibattito alla Tavola Alta dell'evoluzione*, Einaudi, Torino 1999.
- [4] Barbieri M., *The organic codes. An Introduction to semantic Biology*, Cambridge University Press 2003.
- [5] Babushok D.V., Ostertag E.M., Kazazian H.H., Jr, *Current topics in genome evolution: Molecular mechanisms of new gene formation*, *Cell. Mol. Life. Sci.* 64 (2007) 542-554.
- [6] Taylor J.S., Raes J., *Duplication and divergence: the evolution of new genes and old ideas*, *Annu. Rev. Genet.* 38 (2004) 615-643.
- [7] Pigliucci M., *Is evolvability evolvable?*, *Nat. Rev. Genet.* 9 (2008) 75-82.
- [8] Mayr E., Provine W.B., *The Evolutionary Synthesis. Perspectives on the Unification of Biology*, Harvard University Press, Cambridge 1980.
- [9] Müller G.B., *Evo-devo: extending the evolutionary synthesis*, *Nat. Rev. Genet.* 8 (2007) 943-949.
- [10] Kapranov P., Willingham A.T., Gingeras T.R., *Genome-wide transcription and the implications for genomic organization*, *Nat. Rev. Genet.* 8 (2007) 413-423.
- [11] Gerstein M.B., Bruce C., Rozowsky J.S., Zheng D., Du J., Korbel J.O., Emanuelsson O., Zhang Z.D., Weissman S., Snyder M., *What is a gene, post-ENCODE? History and updated definition*, *Genome Res.* 17 (2007) 669-681.
- [12] Noble D., *The Music of Life. Biology Beyond the Genome.*, Oxford University Press, Oxford 2006.
- [13] Jablonka E., Lamb M., *L'evoluzione in quattro dimensioni. Variazione genetica, epigenetica, comportamentale e simbolica nella storia della vita.*, UTET, Torino 2007.
- [14] Ptashne M., *Regolazione genica*, Zanichelli, Bologna 2006.