

# La meccanica statistica affronta i problemi dell'evoluzione

**Luca Peliti**

Dipartimento di Scienze Fisiche, Università "Federico II", Napoli

26 Maggio 2008

**Sommario.** Lo sviluppo dell'interesse per i problemi dell'evoluzione da parte dei cultori di meccanica statistica è favorito, a mio avviso, dallo sviluppo della teoria dei sistemi disordinati e dall'affermarsi del punto di vista di Kimura nella descrizione dei processi evolutivi.

Tra l'11 e il 13 Maggio si è tenuto a Rutgers uno dei convegni semestrali dei cultori USA di Meccanica Statistica. [1] In questo convegno possiamo notare i seguenti contributi:

- C. Callan, Princeton University: Stochastic dynamics of the evolution of gene regulation: simple theory meets real-world data (and survives);
- E. Siggia, Rockefeller University: Predicting the Course of Evolution;
- D. Nelson, Harvard University: Gene Surfing in Microorganisms;
- G. West, Santa Fe Institute: Growth, Innovation and the Pace of Life from Cells to Cities.

Questi contributi affrontano problemi specificamente evolucionistici, ma sono dovuti a scienziati che hanno una formazione di fisica teorica, anzi sono fra i fisici teorici più noti, nel campo della teoria delle particelle elementari (Callan, West) o di quella della materia condensata (Nelson, Siggia). Ovviamente, è comprensibile che scienziati di questo valore si occupino anche di problemi che non rientrano strettamente nel loro campo. Ma è sorprendente che essi ne discutano *nell'ambito di un convegno di Fisica Teorica*.

Qualche anno fa, Mario Ageno formulava così la contrapposizione fra fisica e biologia, separate, come osservava, "da una 'terra di nessuno' inesplorata." La descrizione del mondo fornita dalla fisica è "una descrizione per concetti, e quindi astratta e astorica, espressa in termini di simmetrie e di leggi, costituita esclusivamente di enunciati di validità universale, rigorosamente deterministica a livello dei fatti oppure a livello delle loro probabilità e per questo organizzata nella forma di sistemi logico-deduttivi." La descrizione fornita dalla biologia è invece "una collezione di fatti

particolari, che in nessun caso consentono di formulare vere generalizzazioni prive di eccezioni, fatti particolari organizzabili solo nel racconto delle loro concatenazioni storiche.” [2] Quindi, se la biologia è una disciplina di natura essenzialmente storica, il compito del fisico sarà solo quello di definire i vincoli fisico-chimici entro cui questa storia si è svolta, mentre toccherà al biologo descrivere il comportamento dei sistemi viventi come risultato di un processo storico che non potrà essere completamente spiegato, perché determinato da circostanze irripetibili, a noi sconosciute.

È da notare che questa contrapposizione fra l'uniformità delle leggi fisiche e la varietà delle forme di vita era stata sottolineata da Darwin a conclusione del suo libro sull'origine delle specie: [3]

Vi ha certamente del grandioso in queste considerazioni sulla vita e sulle varie facoltà di essa, che furono in origine impresse dal Creatore in poche forme od anche in una sola; e nel pensare che, mentre il nostro pianeta si aggirò nella sua orbita, obbedendo alla legge immutabile della gravità, si svilupparono da un principio tanto semplice, e si sviluppano ancora infinite forme, vieppiù belle e meravigliose.

Viceversa R. Fisher pensava di trovare nel teorema fondamentale della selezione naturale, secondo il quale l'adattamento (*fitness*) di una popolazione tende a crescere nel tempo, una *regolarità biologica* di valore simile alla regolarità fisica espressa dalla seconda legge della termodinamica: [4, pag. 36]

Si noterà che il teorema fondamentale dimostrato sopra possiede una notevole somiglianza con la seconda legge della termodinamica. Entrambi sono proprietà di popolazioni, o aggregati, vere indipendentemente dalla natura delle unità che li compongono; entrambi sono delle leggi statistiche; ciascuno stipula l'aumento costante di una quantità misurabile, in un caso l'entropia di un sistema fisico, e nell'altro la *fitness*, misurata da  $m$ , di una popolazione biologica.

Tuttavia, egli faceva subito notare le profonde differenze fra le due leggi:

(1) I sistemi considerati in termodinamica sono permanenti; invece, le specie sono soggette ad estinzione, sebbene ci si debba aspettare che il miglioramento biologico abbia luogo fino alla fine della loro esistenza. (2) La *fitness*, sebbene misurata da un metodo uniforme, è qualitativamente differente per ogni organismo differente, mentre si ritiene che l'entropia, come la temperatura, abbia lo stesso significato per ogni sistema fisico. (3) La *fitness* può essere aumentata o diminuita da cambiamenti dell'ambiente, senza che possa reagire quantitativamente su quell'ambiente. (4) Le variazioni d'entropia sono eccezionali nel mondo fisico, perché sono irreversibili, mentre cambiamenti evolutivi irreversibili non costituiscono un'eccezione nel mondo biologico. E infine, (5), i cambiamenti d'entropia portano a una disorganizzazione progressiva del mondo fisico, almeno dal punto di vista umano dell'uso

dell'energia, mentre i cambiamenti evolutivi vengono generalmente riconosciuti come produttori di un'organizzazione sempre più elevata del mondo organico.

Quest'ultima osservazione è riassunta dal biomatematico K. Sigmund nella frase: "Qui progresso, là degradazione: quasi troppo bello per essere vero." [5]

Evidentemente le cose non stanno *esattamente* così. È chiaro che, per la maggior parte dei biologi, vale il detto di Dobzhanski "Niente ha senso in biologia se non alla luce dell'evoluzione." [6] E tuttavia l'evoluzione non corrisponde necessariamente a un incremento costante dell'adattamento delle specie. Da una parte, paleontologi come Stephen Jay Gould e Niles Eldredge sottolineano (riprendendo delle osservazioni di G. Gaylord Simpson [8]) che l'evoluzione appare procedere a singhiozzi, alternando fasi di stasi a improvvise "esplosioni evolutive" (beninteso su scala geologica). [7] Questo punto di vista è noto sotto il nome, non particolarmente felice, di "Teoria degli equilibri punteggiati." Dall'altra parte, una scuola di pensiero, che ha in Motoo Kimura [9] il suo nume tutelare, sottolinea come gran parte delle modifiche che si accumulano nel tempo nelle specie viventi non hanno un effetto rilevante sulla loro *fitness*: sono neutre o solo impercettibilmente dannose. In entrambi i casi, la natura profondamente casuale del processo evolutivo viene ad essere messa sotto la luce dei riflettori. Tuttavia, queste due visioni portano a delle conseguenze importanti sulla natura stessa della teoria evolutiva. Da una parte, c'è la tendenza a sottolineare l'irregolarità e l'imprevedibilità del processo evolutivo, in cui la prevalenza di una forma sull'altra viene in ultima analisi ascritta a circostanze irripetibili: se il grosso dell'innovazione evolutiva ha luogo in (relativamente) pochi episodi nodali, è difficile trovarne una spiegazione differente dal dire "È andata così". Dall'altra, proprio l'esistenza di un "rumore evolutivo", che ha scarsa influenza sulla *fitness*, dà la speranza di vedere emergere delle regolarità statistiche, osservabili e studiabili.

L'interesse che molti cultori della meccanica statistica portano alla teoria dell'evoluzione è dovuto, secondo me, tanto all'affermarsi del punto di vista di Kimura nello studio della struttura del genoma dal punto di vista evolutivo, quanto a un mutamento di paradigma nell'ambito della meccanica statistica stessa.

Il fenomeno che ha accompagnato la nascita della "Teoria neutralista dell'evoluzione" è stata la crescita sempre più rapida dei dati di biologia molecolare a disposizione, e in particolare di sequenze di DNA sempre più lunghe—fino a raggiungere la lunghezza totale del genoma di diverse specie, fra cui *Homo Sapiens*. Questi dati, fin dall'inizio, mostravano una grande e insospettata variabilità individuale, che non poteva essere riconciliata con la teoria corrente se non supponendo che buona parte di questa variabilità non avesse effetti importanti sulla *fitness*. Uno dei risultati più sconcertanti della teoria neutralista, cioè il concetto di "orologio molecolare" (secondo il quale il ritmo con cui si accumulano le sostituzioni in un lignaggio è indipendente dalla taglia della popolazione interessata) è diventato uno strumento fondamentale per ordinare filogeneticamente le sequenze delle varie forme di vita. Questo, ed altri risultati, hanno dato diritto di cittadinanza al punto di vista di Kimura, almeno come punto di partenza per la descrizione dello sfondo sottostante ai processi di adattamento.

Ora, la teoria di Kimura ha un'interpretazione formalmente assai diretta in meccanica statistica.

Un sistema trattato in meccanica statistica è costituito da un numero molto grande di unità, ciascuna delle quali ha una struttura abbastanza semplice, così che il suo stato può essere descritto da un numero limitato di variabili: al limite, da una sola variabile  $\sigma$ , lo "spin", che può assumere due valori:  $\pm 1$ . Queste unità interagiscono fra di loro, secondo delle modalità che sono riassunte da una funzione  $H(\{\sigma\})$  (detta hamiltoniana) della collezione  $\{\sigma\}$  di *tutti* gli stati di ciascuna delle unità che compongono il sistema. A causa di questa interazione, lo stato delle unità cambia molto rapidamente nel tempo, in maniera sostanzialmente incontrollata. Se non è possibile prevedere nei dettagli l'evoluzione dello stato di una qualunque unità, la meccanica statistica può però prevedere, entro certi limiti, le proprietà globali del sistema considerato: lo stato di equilibrio di un sistema è ben descritto assumendo che la probabilità della configurazione descritta dalla collezione  $\{\sigma\}$  degli stati delle singole unità è dato dalla distribuzione di Boltzmann

$$P(\{\sigma\}) \propto e^{-H(\{\sigma\})/k_B T}, \quad (1)$$

dove  $k_B$  è la costante di Boltzmann e  $T$  è la temperatura assoluta (misurata a partire dallo zero assoluto).

È una conseguenza di questo risultato, e del grande numero di unità che compongono i sistemi usuali trattati in meccanica statistica, che essa è in grado di fare predizioni sostanzialmente deterministiche, nonostante la sua natura probabilistica. Consideriamo una quantità osservabile, il cui valore dipende dalla configurazione  $\{\sigma\}$  del sistema:  $A = A(\{\sigma\})$ . Secondo il calcolo delle probabilità, il valor medio  $\langle A \rangle$  di  $A$  nello stato di equilibrio del sistema alla temperatura  $T$  è dato da

$$\langle A \rangle = \sum_{\{\sigma\}} A(\{\sigma\}) P(\{\sigma\}). \quad (2)$$

A causa del grande numero di unità che compongono il sistema, per tutta una certa classe di osservabili il cui valore dipende lentamente dalla configurazione  $\{\sigma\}$  (per cui esso non cambia se un numero limitato di unità cambia stato), il valore *effettivo* assunto da  $A$  sarà praticamente uguale al suo valor medio  $\langle A \rangle$ . Questa è una conseguenza della legge dei grandi numeri: la media della somma di quantità aleatorie indipendenti è pari alla somma delle medie, mentre lo scarto quadratico medio corrispondente tende ad aumentare solo come la radice del numero di addendi. Per osservarne le fluttuazioni è necessario osservare sistemi piccoli, come le particelle browniane considerate da Einstein in uno dei suoi famosi lavori dell'anno mirabile 1905. Quindi: fluttuazioni microscopiche  $\rightarrow$  regolarità macroscopica. Questo stesso meccanismo sottende il teorema fondamentale della selezione naturale di Fisher, menzionato prima: esso è valido solo se la popolazione considerata è tanto grande che il numero effettivo dei discendenti di ciascun tipo di individui è praticamente uguale al corrispondente valore atteso.

D'altra parte, la teoria di Kimura pone l'accento sugli effetti della taglia finita delle popolazioni naturali. In questo caso è possibile, anzi altamente probabile, che

una variante che possiede una *fitness* più elevata del tipo maggiormente rappresentato nella popolazione (il *wild type*) vada persa a causa delle *fluttuazioni di campionamento*. Consideriamo una popolazione di  $N$  individui che si riproduce in maniera asessuata, e supponiamo che a una data generazione la *fitness* di tutti gli individui sia uguale. Sebbene ciascun individuo abbia *in media* un discendente diretto, ci saranno alcuni che ne avranno più d'uno, e una certa parte (una frazione  $1/e$ ) non ne avrà nessuno. Si può far vedere che tipicamente, dopo  $N$  generazioni, tutta la popolazione discenderà da un solo individuo della popolazione originale. Quindi è altamente probabile che alcune caratteristiche che erano presenti nella popolazione originale vengano perse semplicemente perché gli individui che le portavano non hanno avuto discendenza. E si può vedere che questo accade anche se queste caratteristiche sono (moderatamente) benefiche. In effetti, Kimura [10] ha mostrato che, se in una popolazione di taglia  $N$  nasce un mutante che presenta, rispetto al *wild type*, una variazione di *fitness* pari a  $\Delta F$  (supposto piccolo), la probabilità che esso *si fissi*, cioè che venga a riempire tutta la popolazione di suoi discendenti, è data da

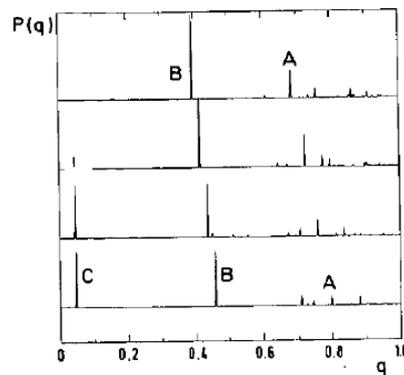
$$P^{\text{fix}}(\Delta F) = \frac{1 - e^{-2\Delta F}}{1 - e^{-2N\Delta F}}. \quad (3)$$

L'andamento di questa espressione è interessante: se  $\Delta F$  è negativo, cioè la mutazione è dannosa,  $P^{\text{fix}}(\Delta F)$  tende a zero come  $e^{-2N\Delta F}$ . Viceversa, se  $\Delta F$  è positiva, cioè la mutazione è benefica, si ha  $P^{\text{fix}}(\Delta F) \simeq \Delta F$ , almeno per piccoli  $\Delta F$ . E per  $\Delta F = 0$  (mutazione neutra),  $P^{\text{fix}}(0) = 1/N$ : poiché il ritmo con cui nascono nuove mutazioni è pari a  $Nu$ , dove  $u$  è la probabilità che nasca una mutazione in un dato individuo, il ritmo con cui le mutazioni si fissano nella popolazione è dato da  $NuP^{\text{fix}}(0) = u$  per mutazioni neutre. Questo è il “risultato sorprendente” a cui avevo accennato prima.

Qualche anno fa ci si è resi conto che le fluttuazioni di campionamento sono analoghe alle fluttuazioni che si osservano nei sistemi disordinati, quelli cioè in cui la hamiltoniana stessa  $H(\{\sigma\})$  è aleatoria. La teoria dei sistemi disordinati è nata al principio degli anni '70, ma si è sviluppata tumultuosamente da quando, nei primi anni '80, Giorgio Parisi [11] ha ottenuto una soluzione presumibilmente esatta e molto sorprendente di un modello paradigmatico: il modello di Sherrington-Kirkpatrick. Qui mi interessa soprattutto notare che da questa soluzione è risultato che le osservabili di un sistema di meccanica statistica disordinato possono essere distinte in due classi: quelle il cui valore, per sistemi grandi, non fluttua percettibilmente (e si comportano così come in un sistema non disordinato) e quelle invece che assumono valori diversi da campione a campione. Consideriamo un sistema di  $N$  unità, in cui lo stato di ciascuna unità è descritto da una variabile  $\sigma = \pm 1$ . Possiamo domandarci quale sia la probabilità che prendendo due configurazioni  $\{\sigma\}$  e  $\{\sigma'\}$  a caso con la distribuzione di Boltzmann (1) esse siano più o meno simili. Una misura della somiglianza di due configurazioni è data dal cosiddetto *overlap* (ricoprimento)  $q(\{\sigma\}, \{\sigma'\})$ , definito da

$$q(\{\sigma\}, \{\sigma'\}) = \frac{1}{N} \sum_i \sigma_i \sigma'_i. \quad (4)$$

Questa quantità è sempre compresa fra  $-1$  e  $+1$ : è uguale a  $+1$  se le due configurazioni sono praticamente identiche (cioè differiscono per un numero di unità trascurabile rispetto a  $N$ ), ed è pari a  $0$  se sono indipendenti l'una dall'altra, per cui esse coincidono in circa metà delle unità e differiscono nelle altre. Ora, se andiamo a valutare la distribuzione degli *overlap* nel modello di Sherrington-Kirkpatrick, per valori di  $N$  molto grandi e per valori di  $T$  abbastanza bassi, otteniamo una distribuzione simile a quella mostrata in figura (la figura contiene *quattro* distribuzioni, che corrispondono a diversi campioni tratti dalla stessa distribuzione). Vediamo una certa popolazione di picchi, per diversi valori di  $q$ . Andando più nei dettagli, ci si può accorgere che questi picchi corrispondono agli *overlap* fra diversi insiemi di configurazioni, e che questi insiemi sono organizzati fra loro in maniera gerarchica. I numerosi piccoli picchi che appaiono per  $q$  prossimo ad  $1$  corrispondono a piccoli insiemi di configurazioni molto simili fra loro. Questi insiemi a loro volta sono raggruppati in insiemi più grandi, tali che due configurazioni che appartengono a diversi piccoli insiemi, ma allo stesso insieme più grande, hanno un *overlap* che corrisponde a uno dei picchi più grandi, relativi a valori più piccoli di  $q$  (come il picco A). E così via. Questa struttura delle configurazioni del modello di Sherrington-Kirkpatrick è chiamata "ultrametricità". Quindi la distribuzione degli *overlap* cambia da campione a campione, e la teoria è in grado di prevedere le proprietà statistiche di questa distribuzione. Adesso posso dirvi che questa figura *non*



**Figura 1.** Distribuzione degli *overlap* nel modello di evoluzione neutra di Derrida-Peliti. I quattro grafici corrispondono a popolazioni discendenti l'una dall'altra, distanziate di un numero fisso di generazioni (dal basso verso l'alto). Da [13].

rappresenta la distribuzione degli *overlap* nel modello di Sherrington-Kirkpatrick, ma piuttosto in un semplice modello di evoluzione neutra introdotto da B. Derrida e dal sottoscritto nel 1991. [12] In questo modello, una popolazione di  $N$  individui si riproduce in maniera asessuata, accumulando via via mutazioni a ritmo costante: e tutti i genotipi hanno la stessa *fitness*. L'*overlap*  $q$  fra due individui è una misura della frazione  $x$  di *loci* che sono differenti nei due individui:  $q = 1 - 2x$ . Poiché le mutazioni si accumulano a ritmo costante, l'*overlap* fra due individui è una misura del tempo trascorso dal loro ultimo antenato comune. Quindi i diversi picchi della distribuzione di  $q$  corrispondono

ad insiemi di coppie di individui che hanno il medesimo ultimo antenato comune. E la struttura ultrametrica della distribuzione corrisponde semplicemente alla struttura ad albero delle genealogie. Metodi abbastanza standard di meccanica statistica permettono di derivare le proprietà statistiche di questa distribuzione: nel caso considerato, per esempio, il valor medio della distribuzione  $P(q)$  è costante nell'intervallo riportato in figura. Evidentemente in questo caso le differenze fra campione e campione non sono per nulla trascurabili.

Più che alla somiglianza fra le strutture della distribuzione di  $q$  in sistemi tanto diversi, l'interesse di questo risultato sta nel mostrare che la teoria dei sistemi disordinati può essere attenta allo stesso tempo alle proprietà generali di una classe di sistemi e alla loro variabilità effettiva, e che in questo modo può affrontare correttamente dei problemi evolutivi. In un lavoro recente, questo punto di vista si è dimostrato utile per analizzare il comportamento evolutivo di certi brevi tratti di DNA che hanno una grande importanza funzionale.

I fattori di trascrizione (*Transcription Factors*, TF) sono delle proteine che si legano a dei siti particolari del DNA, a monte di un gene che codifica una proteina, e ne regolano (positivamente o negativamente) la trascrizione. Spesso accade che uno stesso TF regoli parecchi geni, ognuno dei quali ha un particolare sito di legame. Questi siti hanno di solito lunghezza assai breve (circa una ventina di coppie di basi). Siti diversi per lo stesso fattore di trascrizione presentano di solito sequenze piuttosto differenti: così che l'identificazione di questi siti nella sequenza del DNA è un difficile problema bioinformatico. Tuttavia è possibile, in certi casi, stimare o addirittura misurare l'energia di legame fra il TF e una particolare sequenza di DNA. Questa energia è ben rappresentata da un modello lineare, in cui l'energia di legame totale  $E(s)$  è data dalla somma di contributi indipendenti, definiti dalla base presente in ciascuna posizione della sequenza  $s$  del sito: [14]

$$E(s) = \sum_{i=1}^L \epsilon_{i s_i}, \quad s_i \in \{\text{A, T, G, C}\}. \quad (5)$$

È stato suggerito da una ventina d'anni che l'attività biologicamente rilevante di un particolare sito di legame sia dovuto alla sua energia di legame  $E(s)$ , e quindi che la relativa *fitness* sia in fin dei conti una funzione di  $E(s)$ . Il problema è di comprovare quest'idea e di utilizzarla per approfondire le nostre conoscenze dei siti di legame.

Poiché i siti di legame sono corti, la probabilità che in uno particolare di essi, a un dato istante, sia presente una mutazione è molto bassa. Quindi possiamo pensare che il meccanismo dell'evoluzione di un determinato sito sia dato da una successione di situazioni in cui un singolo mutante si confronta con il *wild type*, e viene fissato con la probabilità  $P^{\text{fix}}$  ottenuta da Kimura (3). Ora supponiamo che la probabilità che una mutazione della sequenza  $s$  produca la sequenza  $s'$  sia data da  $W(s \rightarrow s')$ . Sotto ipotesi abbastanza deboli si può mostrare che la probabilità di trovare la sequenza  $s$  in

un determinato istante in un lungo processo evolutivo neturo è data da una funzione  $P_0(s)$  (la distribuzione di *background*) che soddisfa la relazione

$$P_0(s)W(s \longrightarrow s') = P_0(s')W(s' \longrightarrow s), \quad \forall s, s'. \quad (6)$$

Recentemente Michael Lässig e i suoi collaboratori dell'Università di Colonia hanno mostrato [15] che, se indichiamo con  $P^{\text{fix}}(s \longrightarrow s')$  la probabilità che la sequenza mutante  $s'$  venga a fissarsi secondo l'espressione di Kimura (3) in una popolazione di *wild type*  $s$ , si ha la relazione

$$\frac{P^{\text{fix}}(s \longrightarrow s')}{P^{\text{fix}}(s' \longrightarrow s)} = \exp [2N (F(s') - F(s))]. \quad (7)$$

Questa relazione somiglia a quella soddisfatta dalla distribuzione di *background* rispetto alle mutazioni, e suggerisce che, se le mutazioni producessero, a partire da una sequenza data  $s$  una sequenza casuale  $s'$ , la distribuzione  $P(s)$  delle sequenze in un lungo processo di sostituzione sarebbe data da

$$P(s) \propto e^{2N F(s)}. \quad (8)$$

Notiamo la somiglianza formale di questa espressione con l'espressione (1) della distribuzione di Boltzmann. La fitness  $F(s)$  corrisponde all'opposto dell'energia; e  $2N$  corrisponde all'inverso della temperatura  $k_B T$ .

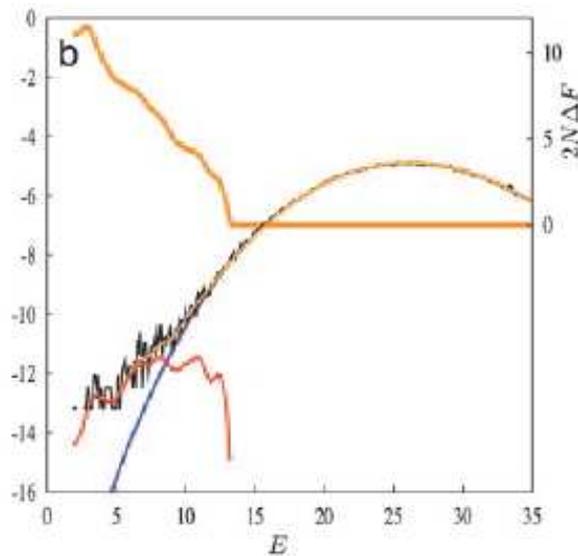
Più precisamente, se teniamo conto della probabilità che la sequenza  $s$  generi la sequenza  $s'$ , otteniamo che la probabilità di osservare la sequenza  $s$  in un sito di legame attivo è data da

$$Q(s) \propto P_0(s) e^{2N F(s)}. \quad (9)$$

Ora, sebbene le sequenze non siano molto lunghe, il numero di sequenze possibili è molto elevato:  $4^L$ , dove  $L$  è la lunghezza del sito. Per  $L = 20$  abbiamo  $2^{40} \simeq 10^{12}$ . Non è possibile sperare di ottenere una buona statistica su questa distribuzione. Però, se la fitness dipende solo dall'energia di legame, possiamo provare a *proiettare* questa distribuzione dallo spazio delle sequenze a quello delle energie di legame, raggruppando insieme tutte le sequenze che hanno lo stesso valore delle energie di legame  $E$ . Otterremo così una distribuzione di *background*  $P_0(E)$  e una distribuzione sotto selezione  $Q(E)$ , tale che

$$\frac{Q(E)}{P_0(E)} = e^{2N F(E)}. \quad (10)$$

Notiamo che queste distribuzioni riguardano la probabilità di osservare una sequenza con energia di legame  $E$  per un particolare sito di legame, durante un lungo processo evolutivo in condizioni fissate. Però possiamo supporre (e verificare a posteriori) che i vari siti di legame relativi a uno stesso TF evolvano secondo processi evolutivi equivalenti: in questo caso è lecito supporre di ottenere la distribuzione  $Q(E)$  *guardando la distribuzione di  $E$  per i vari siti di legame dello stesso TF lungo il genoma*. Con questa ipotesi è possibile valutare  $P_0(E)$  valutando la distribuzione di  $E$  lungo tutto il genoma non codificante dell'organismo considerato, e ricavare  $F(E)$  dal confronto



**Figura 2.** Distribuzione delle energie di legame per il TF CRP in *Escherichia coli*, su scala logaritmica. Linea nera: distribuzione osservata. Curva arancione: distribuzione di *background*. Linea rossa: Distribuzione in eccesso:  $Q(E) = P^{\text{tot}}(E) - Q(E)$ . La linea arancione in alto (asse di destra) rappresenta la stima di  $2N \Delta F$  ottenuta dalla relazione (10). Da [15].

fra la distribuzione  $Q(E)$  e la distribuzione di *background*. Queste distribuzioni sono abbastanza ben campionate, perché il numero di siti di legame per alcuni TF è dell'ordine del centinaio, mentre il numero di sequenze che forniscono la distribuzione di *background* è pari alla lunghezza del genoma non codificante diviso la lunghezza del sito di legame, ed è quindi molto elevato.

Questo programma è stato portato a termine per il TF “cAMP response protein” (CRP) in *Escherichia coli* e, più recentemente, per il TF Abf1p in *Saccharomyces cerevisiae*, come comunicato da C. Callan nel convegno di Rutgers.

Quando si era appena incominciato ad affrontare i temi dell'evoluzione dal punto di vista della meccanica statistica, ci sono stati a volte dei lavori in cui l'ambizione si coniugava a una grande generalità—e a un'altrettanto scarsa concretezza. Ne è un esempio il “modello semplice di evoluzione” dovuto a Bak e Sneppen, [16] che forniva una delle molte spiegazioni degli equilibri punteggiati sulla base di una semplicissima rappresentazione della coevoluzione. I lavori più recenti tendono maggiormente alla concretezza, confortati in ciò dalla disponibilità delle sequenze di un notevole numero di specie, e dei risultati degli esperimenti di evoluzione condotti sui virus, sui batteri e su alcune piante. Mi sembra che sia passato di moda l'atteggiamento riduzionistico testimoniato per esempio dai lavori di Lotka, [17] che considerava la selezione naturale come un principio fisico, corrispondente alla tendenza a massimizzare il flusso d'energia. La possibilità di investigare la complessità e la variabilità mediante la teoria dei sistemi disordinati permetta un approccio più modesto e più concreto, che potrà portare a risultati importanti tanto per la biologia che per la fisica teorica.

**Bibliografia**

- [1] 99th Statistical Mechanics Conference, Rutgers University, Maggio 2008.
- [2] M. Ageno, *Dal non vivente al vivente* (Roma-Napoli: Theoria, 1991).
- [3] C. Darwin, *Sulla origine delle specie per elezione naturale, ovvero conservazione delle razze perfezionate nella lotta per l'esistenza*, trad. G. Canestrini, e-book Liber Liber.
- [4] R. A. Fisher, *The genetical theory of natural selection* (Oxford: Clarendon Press, 1930). Online in Open Library. La traduzione è mia.
- [5] K. Sigmund, *Games of Life* (Oxford: Oxford U.P., 1994).
- [6] T. Dobzhanski, *American Biology Teacher* **35**, 125–129 (1973).
- [7] N. Eldredge and S.J. Gould, in: T.J.M. Schopf (ed.), *Models in Paleontology* (San Francisco: Freeman Cooper, 1973), pagg. 82–115.
- [8] G. G. Simpson, *Tempo and mode in evolution* (New York: Columbia U. P., 1944).
- [9] M. Kimura, *The neutral theory of molecular evolution* (Cambridge: Cambridge University Press, 1983).
- [10] M. Kimura, *Genetics* **47**, 713–719 (1962).
- [11] Vedi M. Mézard, G. Parisi e M.A. Virasoro, *Spin glass theory and beyond* (Singapore: World Scientific, 1987) per una raccolta degli articoli originali, e per le prime applicazioni biologiche.
- [12] B. Derrida e L. Peliti, *Bull. Math. Biol.* **53** 355–382 (1991).
- [13] P.G. Higgs e B. Derrida, *J. Phys. A: Math. Gen.* **24**, L985–L991 (1991).
- [14] O. Berg e P. von Hippel, *J. Mol. Biol.* **284**, 723–750 (1987); G. D. Stormo e D.S. Fields, *Trends Biochem. Sci.* **23**, 109–113 (1998).
- [15] V. Mustonen e M. Lässig, *Proc. Natl. Acad. Sci. USA* **102**, 15936–15941 (2005).
- [16] P. Bak e K. Sneppen, *Phys. Rev. Lett.* **71**, 4083–4086 (1993).
- [17] A.J. Lotka, *Proc. Natl. Acad. Sci. USA* **8**, 147–151 (1922); *ibid.* 151–154 (1922).