
BOLLETTINO

UNIONE MATEMATICA ITALIANA

Sezione A – La Matematica nella Società e nella Cultura

EDOARDO BONCINELLI

Il genoma come testo in attesa di decifrazione

Bollettino dell'Unione Matematica Italiana, Serie 8, Vol. 7-A—La Matematica nella Società e nella Cultura (2004), n.2, p. 237–247.

Unione Matematica Italiana

http://www.bdim.eu/item?id=BUMI_2004_8_7A_2_237_0

L'utilizzo e la stampa di questo documento digitale è consentito liberamente per motivi di ricerca e studio. Non è consentito l'utilizzo dello stesso per motivi commerciali. Tutte le copie di questo documento devono riportare questo avvertimento.

*Articolo digitalizzato nel quadro del programma
bdim (Biblioteca Digitale Italiana di Matematica)
SIMAI & UMI*

<http://www.bdim.eu/>

Il genoma come testo in attesa di decifrazione (*).

EDOARDO BONCINELLI

Si dice *genoma* il patrimonio genetico di una specie, cioè il complesso delle informazioni biologiche necessarie per la sopravvivenza degli organismi appartenenti a tale specie. Così, nel nostro genoma sono contenute le informazioni per costruire un essere umano, per tenerlo in vita e per farlo riprodurre. Queste istruzioni prendono la forma di un gigantesco testo lineare, scritto utilizzando un alfabeto contenente soltanto quattro lettere, dette più propriamente *nucleotidi* — A, G, C e T — e portato da un'enorme molecola di DNA (acido desossiribonucleico) che si presenta praticamente unidimensionale. È lunga infatti un paio di metri e ha uno spessore inferiore a un nanometro. Nel caso del nostro proprio genoma il testo consta di circa tre miliardi di nucleotidi (equivalente a poco meno di 1 Gigabyte di informazione) che si succedono uno dietro l'altro senza interruzioni né alcun altro tipo di segnale. Eppure, la sostituzione di uno solo di questi nucleotidi può fare la differenza fra un uomo vivo e un uomo morto.

Ciascuna molecola di DNA è costituita in realtà da una coppia di filamenti che corrono antiparalleli e che portano la stessa informazione genetica. Ciascun tratto dei due filamenti contiene infatti la stessa informazione scritta in maniera complementare: laddove sull'uno c'è una A, sull'altro c'è una T e viceversa; laddove sull'uno c'è una G, sull'altro c'è una C e viceversa. Questa particolarità, scoperta nel 1953 da Watson e Crick, garantisce stabilità all'informazione genetica, facilita enormemente il processo di duplicazione della molecola stessa — che ha luogo quotidianamente, ad ogni divisione cellulare — e ne assicura una certa uniformità. Visto dal di fuori, infatti, ogni tratto del genoma è essenzialmente uguale ad ogni altro. Quando parlo del

(*) Conferenza tenuta per invito in occasione dell'Assemblea Generale U.M.I. del 22 maggio 2004.

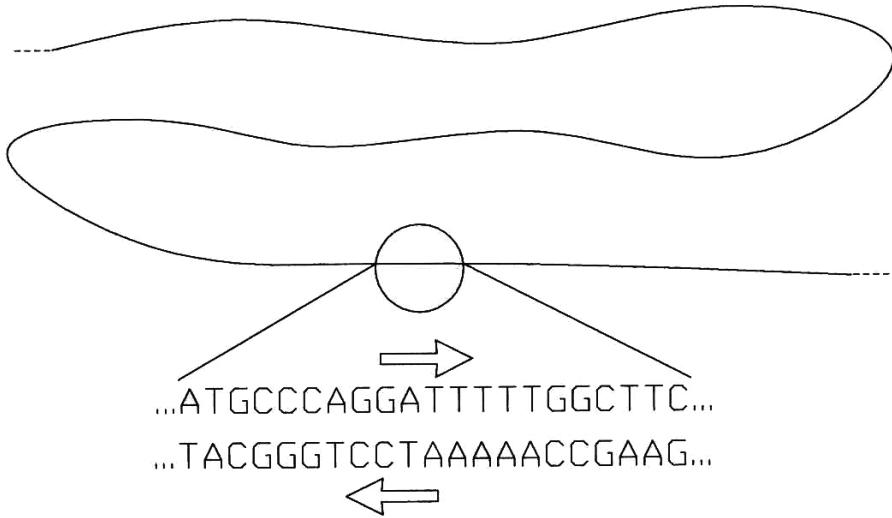


Fig. 1. – Rappresentazione schematica della gigantesca molecola di DNA che porta l'informazione genetica. Si tratta di un testo scritto senza interruzione con un alfabeto di sole quattro lettere: A, G, C e T. Un piccolo tratto è mostrato ingrandito. Consta in realtà di un doppio filamento. I due filamenti sono orientati e antiparalleli e portano la stessa informazione genetica codificata in maniera complementare. Per ogni tratto viene «letto» dalla cellula solo uno dei due filamenti, ad esempio quello che porta la sequenza ... ATGCCCAGGATTTTTGGCTTC ...

genoma come testo mi riferisco a quello portato da uno solo dei due filamenti e nel seguito non menzionerò più l'esistenza di un doppio filamento. Ciò è accettabile, dal momento che l'informazione contenuta in ciascuno dei due filamenti è esattamente la stessa.

Una copia del genoma è contenuta in ognuna delle nostre cellule. Questa è una particolarità interessante degli esseri viventi: essi possiedono una copia del programma «patrimonio genetico» per ciascuna delle proprie cellule, che nel caso della nostra specie sono dell'ordine di 10^{13} . Nei robot, servomeccanismi o calcolatori progettati dall'uomo si trovano in genere una copia o due del programma logico, mentre in un organismo vivente ce ne sono migliaia e migliaia. Può darsi che ciò non sia rilevante, ma può anche essere che abbia una grande importanza che al momento ci sfugge. Una cosa comunque

non ci sfugge. Il fatto che ogni cellula abbia una sua copia del genoma rende possibili le trasfusioni di sangue, i trapianti di midollo osseo e di molti organi. Se le diverse cellule perdessero infatti «la bussola», cioè il loro corredo di codici di comportamento, appena separate dal corpo al quale appartengono, nessuna di queste procedure avrebbe alcun senso. Le cellule espiantate vivrebbero «di rendita» per qualche ora o qualche giorno sulla scorta delle istruzioni ricevute prima del prelievo e poi degenererebbero. Ciò non accade appunto perché ogni cellula porta con sé il proprio programma di lavoro.

TTT	Phe	TCT	Ser	TAT	Tyr	TGT	Cys
TTC	Phe	TCC	Ser	TAC	Tyr	TGC	Cys
TTA	Leu	TCA	Ser	TAA	STOP	TGA	STOP
TTG	Leu	TCG	Ser	TAG	STOP	TGG	Trp
CTT	Leu	CCT	Pro	CAT	His	CGT	Arg
CTC	Leu	CCC	Pro	CAC	His	CGC	Arg
CTA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CTG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
ATT	Ile	ACT	Thr	AAT	Asn	AGT	Ser
ATC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
ATA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
ATG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GTT	Val	GCT	Ala	GAT	Asp	GGT	Gly
GTC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GTA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GTG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Fig. 2. – Il codice genetico.

Legenda: Ad ogni terna (detta anche *tripletta* o *codone*) di nucleotidi contigui corrisponde uno dei 20 aminoacidi naturali:

Ala = alanina; Arg = arginina; Asn = asparagina; Asp = acido aspartico; Cys = cisteina; Gly = glicina; Gln = glutammina; Glu = acido glutammico; His = istidina; Ile = isoleucina; Leu = leucina; Lys = lisina; Met = metionina; Phe = fenilalanina; Pro = prolina; Ser = serina; STOP = tripletta di terminazione; Thr = treonina; Trp = triptofano; Tyr = tirosina; Val = valina.

Per nostra fortuna, il testo che definisce il genoma può essere mentalmente suddiviso in migliaia di capitoli di senso compiuto che siamo soliti chiamare *geni*. Ogni gene porta l'informazione per almeno una funzione biologica, specificando in genere la struttura di una catena proteica sulla scorta di un codice universale predeterminato, detto *codice genetico*, che ad ogni terna di nucleotidi contigui presenti sul DNA fa corrispondere un determinato aminoacido nella catena proteica. Ad esempio ATG significa metionina, TTT significa fenilalanina, TAA significa fine della catena proteica.

Storicamente sono venuti prima i geni e il loro studio e poi, molto di recente, si è passati alla determinazione e allo studio del genoma. Molti fra i primi geni sono stati individuati perché un'alterazione, che noi chiamiamo *mutazione*, di uno o più nucleotidi nel loro proprio frammento di testo può comportare l'alterazione di una caratteristica fisica dell'organismo che la porta, alterazione che può giungere fino alla morte dell'organismo stesso. Il rapporto biunivoco fra alterazioni di struttura di un gene di questo tipo, codificante caratteri cosiddetti *monofattoriali*, e alterazioni somatiche presenti nell'individuo che le porta ha facilitato enormemente l'individuazione dei primi geni. Stiamo parlando ovviamente sulla base del senno di poi. Lo sforzo che ha portato a individuare i primi geni è stato in realtà titanico. Basti pensare che nel 1968, quando sono entrato io in biologia, i geni umani noti erano 5 o 6.

Un gene dietro l'altro ne sono stati identificati moltissimi e di ciascuno di loro è stata determinata in dettaglio l'effettiva sequenza nucleotidica. Il mio gruppo di ricerca ne ha identificati complessivamente una sessantina e ne ha studiato il ruolo e il funzionamento. Ogni laboratorio faceva lo stesso sui geni che più gli interessavano, fino a che non ci si è resi conto che forse conveniva determinare una volta per tutte la sequenza nucleotidica completa dell'intero genoma. Da quel punto in poi chiunque fosse interessato al gene X, se lo andava a cercare nella banca dati e se lo studiava partendo da una sequenza contenuta *in silico*, cioè in un computer, invece di determinarsela di volta in volta con procedure *in vitro*, o compito ancora più impegnativo, *in vivo*. L'ultimo decennio del XX secolo è coinciso con

l'avvio, l'esecuzione e la conclusione del cosiddetto Progetto Genoma che ha portato appunto alla determinazione della sequenza nucleotidica pressoché completa del nostro genoma, e di quello di vari altri organismi di un certo interesse.

La conclusione di questa impresa si presenta allo stesso tempo come il coronamento di decenni di ricerca biologica e come l'inizio di una nuova era della medicina e della biologia, un'era che è già stata definita *post-genomica*. Tale innegabile clamoroso successo è equivaleso però soltanto all'acquisto di un'enciclopedia. Adesso questa va letta e studiata voce per voce. Il vero problema è che per il momento molte voci non le sappiamo proprio leggere. Abbiamo davanti a noi l'intera sequenza nucleotidica del genoma e la possiamo consultare in ogni momento, ovviamente con l'aiuto del computer. Occorrerebbero infatti 80 anni per poterla leggere ad alta voce. Ne sappiamo leggere al momento un 3% circa, quella frazione che corrisponde chiaramente alla sequenza di una moltitudine di catene proteiche codificate sulla scorta delle corrispondenze del codice genetico. Sul resto possiamo fare solo congetture, anche se appoggiate alle conoscenze biologiche di cui disponiamo.

Che cosa ci possiamo aspettare di trovare nel genoma? Possiamo ipotizzare l'esistenza di almeno tre compartimenti ideali che contengano, rispettivamente, i geni di cui già conosciamo qualcosa, quelli dei quali non sappiamo quasi niente ma dei quali riusciamo ad immaginare l'esistenza e infine quelli la cui natura ci è al momento ignota. Il primo compartimento è quello dei geni corrispondenti ai caratteri monofattoriali, quelli cioè che comportano una corrispondenza diretta fra una mutazione e una caratteristica biologica - normale, quasi normale o patologica. Non conosciamo ancora tutti quanti i geni di questo tipo, ma ne conosciamo moltissimi e soprattutto ne possediamo la chiave. Sappiamo ricostruire la sequenza della catena proteica che ciascuno di questi specifica, possiamo studiarne l'azione e talvolta anche il ruolo nell'economia complessiva del corpo. Appartengono a questo compartimento anche molti geni dei quali non abbiamo per il momento una conoscenza diretta nella nostra specie, ma che sono presenti in specie animali affini o che presentano qualche somiglianza con geni conosciuti. Tutto fa pensare quindi che esistano da qualche parte del nostro ge-

noma. A questi vanno aggiunti anche tutti quei geni ai quali non corrisponde alcuna alterazione biologica nota al momento, ma che sono stati identificati come responsabili di alcune funzioni biologiche fondamentali, quali la duplicazione cellulare, la comunicazione fra le cellule, la respirazione, la digestione e via discorrendo.

Osserviamo di passaggio a questo proposito che i nostri geni non sono lì per portarci sventure e malanni. I geni esistono al contrario per farci stare bene, per farci sviluppare e condurre una vita normale. Qualcuno di questi qualche volta non funziona a dovere e causa una malattia o ci predispone a qualche altra. Insomma, il nostro genoma non è un repertorio di affezioni e di disgrazie, ma un insieme di funzioni vitali, necessarie e sufficienti per farci vivere una vita sana e attiva. È la prospettiva storica che ci porta a nominare molti di questi geni sulla base delle malattie che possono causare: il gene dell'albinismo, quello della fibrosi cistica, quello della distrofia muscolare, quello del tumore al seno e via discorrendo. Questi geni sono stati infatti individuati nel passato partendo dall'osservazione dei danni causati da una loro alterazione. A quell'epoca non sarebbe stato possibile identificare un gene che non fosse alla base di qualche patologia, meglio se ereditaria. Oggi non è più così e i geni si possono studiare nella loro fisiologia e non soltanto nella loro patologia.

Non tutti i difetti genetici e non tutte le nostre caratteristiche biologiche, patologiche o meno, dipendono dall'azione di un gene singolo. Anzi, la stragrande maggioranza delle nostre caratteristiche personali — dall'altezza alla forma del naso e degli orecchi, dal ritmo cardiaco alla pressione sanguigna, dall'intelligenza alla perseveranza — dipendono dall'azione combinata di molti geni — cinque, cinquanta, cinquecento o anche più — e dalla loro interazione con l'ambiente in cui viviamo. Esistono anche malattie molto diffuse — il diabete, l'ipertensione, il ritardo mentale, la labilità psicologica, l'emigrania, l'epilessia, la disposizione alle malattie cardiocircolatorie, allo sviluppo di allergie di vario tipo, al decadimento senile e ai vari tipi di tumori — che dipendono dal mancato funzionamento o da un deficit parziale di molti geni diversi. Della stragrande maggioranza di queste non si conosce ancora né il numero né la natura dei geni responsabili. Il secondo capitolo della saga del genoma riguarderà

appunto l'individuazione di quanti e quali geni cooperano alla determinazione di ciascuno di questi tratti biologici, detti *multifattoriali* per distinguerli da quelli più propriamente monofattoriali. Non sarà un'impresa semplice. Richiederà anzi molto impegno e una grande inventiva. Per studiare questa enorme massa di geni che agiscono di concerto non si potranno ripetere le medesime operazioni che sono state messe a punto per i geni del primo tipo. Sono troppi e ciascuno di loro esplica un'azione molto sottile.

Il progetto implicante la definizione della base genetica dei tratti multifattoriali è solo all'inizio ma presto si svilupperà fino a dare frutti che riusciamo a mala pena ad immaginare. Si tratterà in sostanza di dare corpo alle ombre, di materializzare l'impalpabile, di nominare ciò che oggi è innominato. Consideriamo anche solo quale esaltante impresa sarà fare l'inventario dei geni che sono alla base del nostro linguaggio e della possibilità di acquisirlo nei primi anni della nostra vita! Non sappiamo assolutamente che cosa troveremo, ma qualcosa troveremo e sarà certamente interessante. Personalmente, aspetto con ansia il momento in cui alcuni di questi problemi cominceranno ad essere risolti e ne sapremo di più, di noi stessi e della nostra particolare visione del mondo.

È bene aggiungere che i due compartimenti genici di cui abbiamo appena parlato non sono mutuamente esclusivi. Niente impedisce ad un gene del primo tipo di comportarsi anche come uno dei geni che determinano collettivamente una caratteristica biologica multifattoriale. Anzi, questa rappresenta forse la regola, perché i due tipi di geni non hanno probabilmente niente di diverso. Per nostra comodità logica siamo indotti a distinguere i due compartimenti sulla base dei loro effetti. Va anche detto che è altamente improbabile che una determinata caratteristica biologica, sia essa normale o patologica, sia determinata esclusivamente e al cento per cento da un solo gene. Nessun gene in sostanza agisce in completo isolamento, anche se per alcuni caratteri questa osservazione è più pertinente che per altri.

Ma non è ancora tutto. È quasi certo che nel nostro genoma non ci sono solo i geni dei tipi di cui abbiamo parlato fino adesso. Ci sarà verosimilmente dell'altro che non riusciamo neppure a immaginare e che per questo al primo impatto neppure sapremo vedere. Ci sa-

ranno altri tipi di geni, organizzati in modo diverso e forse anche scritti in modo diverso. Quello che è certo è che esistono vaste regioni del genoma che non contengono sequenze nucleotidiche che possano essere facilmente interpretate come specificanti la struttura di alcuna catena proteica, per esempio perché sono piene di quelle particolari terne di nucleotidi contigui che indicano la terminazione della catena proteica, indicate come STOP nella Figura 2. Non è affatto impossibile che queste regioni che a noi appaiono oggi prive di geni siano in realtà piene di qualcos'altro che non conosciamo. Sembra anzi molto probabile. Possiamo speculare che quel qualcosa in più sia necessario per mettere in atto una regia di livello superiore degli avvenimenti che hanno luogo durante lo sviluppo embrionale oppure nel processo di codificazione e/o di recupero dei ricordi. È molto probabile che alla distanza questa si riveli la parte più interessante della saga della decifrazione del nostro genoma, cioè della nostra natura e delle nostre origini.

Che cosa possiamo pensare di questi geni «non canonici»? Possiamo pensare che specificino la struttura di molecole di RNA, perché si sa anche oggi che alcuni tratti di DNA lo fanno. Possiamo pensare che specificino la sequenza di proteine con un codice a due lettere invece che a tre, oppure a quattro lettere, oppure con un codice «a salto». Oppure qualcosa di ancora diverso. Quello che è certo è che ci troviamo dinanzi ad un lunghissimo testo che è ancora quasi tutto da interpretare. Occorreranno, chissà?, biologi, informatici, matematici, logici, medici, fisici o chimici per trovarne le chiavi di lettura. O più probabilmente una sapiente combinazione di queste diverse competenze. Occorrerà certamente disporre di calcolatori e di software sempre più avanzati e anche una ... scintilla d'ingegno. Il compito è chiaro, la via da percorrere molto meno. Ne vedremo delle belle.

Potremmo chiederci a questo punto quale può essere il ruolo del matematico in questa impresa. È difficile dirlo. Si tratta di un procedimento di decifrazione per il quale non si sa neppure da che parte cominciare. In una circostanza del genere l'intuizione dice che la matematica potrebbe essere di valido aiuto in considerazione della sua grande capacità di astrazione e della propensione a concentrarsi sull'essenza dei fenomeni. La mia impressione è che occorra trovare

una via interpretativa partendo da un numero ridottissimo, se non nullo, di ipotesi. Potrebbe darsi cioè che alcune regioni del genoma debbano essere lette prendendo due nucleotidi in una localizzazione seguiti da tre che si trovano 40 nucleotidi a valle e da altri due che si trovano 15 nucleotidi ancora a valle, se non a monte, e così via. Non si può escludere nemmeno che la forma tridimensionale che la molecola di DNA assume all'interno di una cellula vivente non fornisca una falsariga. Il punto è che oggi non si sa, e non è detto che sia facile rendersene conto sperimentalmente. In questa ottica, se la matematica o qualsiasi altra disciplina fornissero una serie di modelli verificabili, il lavoro sperimentale sarebbe enormemente agevolato. In fondo questo è per me il ruolo fondamentale di ogni approccio teorico e di ogni modellizzazione.

Personalmente, non ho mai visto come rilevante l'apporto che la matematica ha dato fino ad oggi alla biologia. Sarà colpa della matematica, al momento inadeguata, o sarà colpa della biologia, intrattabile, sta di fatto che la cosiddetta biologia matematica non è una parte significativa della biologia d'oggi. Serve al massimo a confermare certe conclusioni, come nel caso della genetica delle popolazioni, o a dare una veste di «scienza esatta» ad alcune branche della biologia che tutto sono fuor che quello. Perché? Azzardo qualche ipotesi.

A differenza di quelli fisici e di parte di quelli chimici, la maggior parte dei fenomeni biologici non sono «necessari» e mostrano anzi tutti i caratteri della contingenza. Questo non significa che gli oggetti biologici non seguano le leggi della fisica e della chimica. Il punto è che queste scienze dettano i «vincoli» entro i quali i fenomeni biologici si muovono e, si direbbe, si sbizzarriscono, ma non «prescrivono» in positivo quello che deve accadere nelle cellule e negli organismi. Per tutti quegli aspetti di vincolo chimico-fisico il trattamento matematico va ovviamente bene, ma non è quella l'essenza della biologia. Il carattere contingente dei fenomeni biologici e il fatto che gli organismi sono arrivati ad essere quello che sono attraverso un percorso erratico, frutto di una combinazione di casualità e di opportunismo nel quadro di un continuo riciclaggio di entità preesistenti, viene talvolta riassunto nell'affermazione che la biologia è una scienza storica.

È inoltre essenzialmente una scienza dei piccoli numeri. In una po-

polazione naturale possono esservi moltissimi individui, ma questo numero molto raramente supera il miliardo e i vari individui sono tutt'altro che identici, anche solo potenzialmente, a parte, forse, i microrganismi unicellulari. Dal canto loro, le cellule di un organismo sono tante, ma appartengono a tanti tipi diversi e con localizzazioni diverse. Ad esclusione delle cosiddette staminali, ogni cellula è unica, è *quella* cellula. La biologia si occupa insomma di quasi tante classi naturali quante sono le cellule presenti sulla Terra, più o meno 10^{30} .

Se poi scendiamo all'interno della cellula, constatiamo che la biologia non è intrinsecamente una scienza quantitativa, ma piuttosto qualitativa. Quando si studia la fisiologia della cellula, che è poi quella della vita, quello che conta non è *quante* molecole sono presenti in un dato momento, ma *quali*. Le quantità non possono ovviamente essere irrilevanti, ma non sono, almeno per quanto possiamo capire oggi, il punto essenziale. Non dimentichiamo che in una cellula ci sono 2, solo 2, molecole di DNA, qualche centinaio di molecole dei vari tipi di RNA messaggero — il portatore dell'informazione genetica dal DNA alle proteine — e qualche migliaio di proteine di vario tipo. Le molecole d'acqua che vi sono contenute esibiranno certamente un comportamento statistico, come quelle dei vari gas in essa disciolti, ma non è questo che fa la cellula e la fa vivere. Quello che colpisce è la grande regolarità di funzionamento, implementata da un numero ridottissimo di agenti materiali, in barba a tutte le considerazioni oggi di moda sulla morte del determinismo e sulla non prevedibilità, se non statistica, dei sistemi dinamici di vario tipo.

Il punto essenziale è che ogni cellula possiede un genoma. Ogni cellula è se stessa *più* il suo genoma, mentre un sasso, o un granello di sabbia, è se stesso e basta. Se la vogliamo mettere in una maniera diversa, la «memoria» del sasso o del granello di sabbia è «azzerrata». Anche essi hanno ovviamente una storia, ma questa è irrilevante. La cellula e tutti gli organismi del mondo portano con sé la propria storia, la propria «memoria», sotto forma del loro genoma, che consultano in continuazione. È una memoria viva, altrettanto viva della realtà momentanea della cellula stessa. Si tratta di un presente inscindibile dal proprio passato, o per meglio dire dalla ricreazione momentanea e specifica del proprio passato. Il non farlo non permettereb-

be alla cellula o all'organismo in questione di passare il controllo di qualità imposto dalle leggi della fisica e della chimica.

Lavorare su numeri piccoli e possedere una memoria storica comportano il fatto che per gli esseri viventi la quantità fisica più rilevante non sia l'energia, ma l'informazione. L'informazione e la sua gestione sono importanti per ogni sistema materiale, ma si trovano al fondamento della biologia, come è stato rilevato fino dagli albori della cosiddetta biologia molecolare, verso la metà del secolo scorso. In tutto questo l'informazione contenuta nel genoma fa la parte del leone e leggerla e interpretarla è indispensabile per la vita di quella data specie. In fondo è proprio il genoma l'elemento biologico più tipico e affidabile. Ed è forse anche il più adatto ad essere studiato con gli strumenti delle varie scienze, a cominciare dalla più antica, la matematica. Magari una matematica delle qualità, chi sa?

Un'ultima osservazione. Sappiamo che un nucleotide su mille circa è diverso fra i genomi di due persone prese a caso. Si tratta di più di un milione di differenze nucleotidiche fra due individui qualsiasi della stessa specie. Non tutte queste differenze genomiche hanno ovviamente un effetto pratico. La stragrande maggioranza non ne ha alcuno; un certo numero di queste comportano differenze di poco conto; un'esigua minoranza causa differenze degne di nota. Ma qual è allora il genoma umano? Quale delle sue forme scegliamo per rappresentare la nostra specie? Se una determinata differenza nucleotidica è irrilevante o meno lo si può stabilire soltanto dopo aver osservato la natura di tutti gli esseri umani e averne determinato l'intera sequenza nucleotidica. Ma chi mi dice che l'individuo che non ho ancora analizzato non avrebbe qualcosa di importante da insegnarmi? Esiste insomma una sorta di Principio di Indeterminazione genomico: posso sapere qual è il genoma di una data specie solo dopo aver analizzato tutti gli individui possibili appartenenti a quella specie; e questo richiede ovviamente un tempo infinito. Poiché il tempo non è infinito, nemmeno potenzialmente, considerando che dopo un po' la specie scompare o si trasforma, mi dovrò accontentare di una conoscenza approssimata di tale genoma. Per gli scopi pratici questo non è assolutamente un dramma, ma pone un'interessante limitazione teorica.