

Analisi del DNA e probabilità

davide.sisti@uniurb.it

1. IL DNA: CAMPO D'INDAGINE

Il DNA è una lunghissima molecola, che sta alla base della chimica degli esseri viventi. Da un punto di vista strutturale, il DNA è una doppia elica, costituita da due filamenti formati dalla diversa alternanza di 4 diverse molecole chiamate nucleotidi: essi sono la timina (T), l'adenina (A), la guanina (G) e la citosina (C). La disposizione in sequenza dei nucleotidi nel DNA di un individuo è la causa della sua unicità, sia che esso sia un organismo unicellulare, una pianta o un animale. Il filamento a doppia elica è, di norma, lunghissimo, mentre il suo diametro è estremamente piccolo; si stima infatti che il DNA umano sia lungo circa due metri, mentre il suo diametro è di solo due milionesimi di metro. Per meglio intuire la piccolezza della misura, se si potesse aumentare il diametro del DNA fino ad un centimetro, una pallina da golf dovrebbe proporzionalmente assumere un diametro pari circa alla distanza Roma – Milano! È singolare pensare che ogni singola cellula (di dimensioni microscopiche) contenga l'intero DNA, strettamente avvolto su se stesso, all'interno del proprio nucleo.

Lo studio del DNA è stato, fino a circa due decenni fa estremamente difficoltoso; ora invece è utilizzata una tecnica rivoluzionaria, chiamata PCR (*polymerase chain reaction*), che permette di amplificare in poco tempo specifici tratti di DNA presenti all'inizio solo in quantità minima. La reazione PCR fu proposta da uno scienziato controverso, Kary Mullis, che nel 1993 vinse il premio Nobel per la chimica. Mullis dichiarò esplicitamente che se non avesse fatto uso di LSD non avrebbe mai potuto immaginare il meccanismo per il quale oggi è possibile la PCR. Mullis stesso racconta:

Non avrei mai immaginato, all'inizio, l'enorme diffusione che questa invenzione avrebbe avuto. Allora io lavoravo alla Cetus e tentavo di amplificare in modo specifico, senza errori, in laboratorio, delle sequenze naturali di Dna. Cominciai a

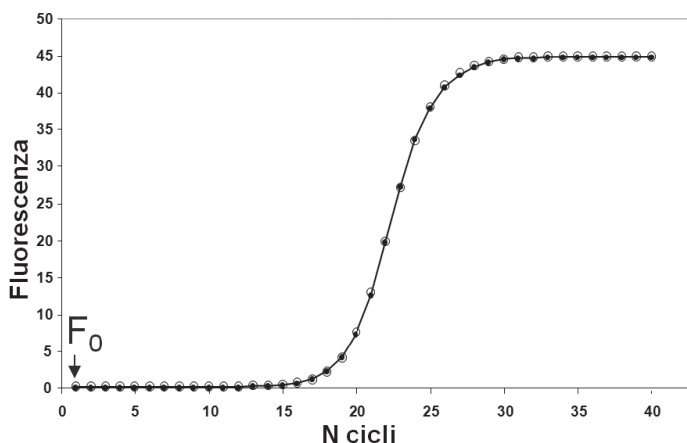
fare copie di copie di copie. Ma era un disastro, in quanto gli errori di sequenza si moltiplicavano senza controllo. Ed era, inoltre, una procedura noiosissima, basata su cicli di raffreddamento e riscaldamento dei prodotti e dell'enzima specifico che duplica il Dna in natura. La sorte volle che, all'altro estremo del corridoio dei laboratori Cetus, lavorasse una collega intenta a studiare le fermentazioni; lavorava con enzimi estratti da batteri che sono detti termofili, perché sopportano temperature assai elevate senza alcun danno biologico. La polimerasi di questi batteri resiste alla temperatura di un bagno molto caldo, mentre, a quella temperatura, il Dna si srotola, si scinde in due eliche distinte e complementari, ciascuna delle quali può darne un'altra e queste darne altre due e così via, appunto, in una reazione a catena. (Piattelli Palmarini 2004)

Oggi eseguire una reazione di PCR è alla portata di qualunque laboratorio; il suo principale limite ad una estesa applicazione è nel costo dei reagenti. Per eseguire una reazione di amplificazione PCR occorre conoscere almeno parte della sequenza di DNA che si vuole amplificare, effettuare una minima purificazione del DNA stesso, sintetizzare in laboratorio dei corti innesci di DNA sulla base della sequenza da amplificare. In pratica tramite la reazione PCR si ottiene un numero elevatissimo (di molto superiore a un miliardo) di molecole di DNA identiche, a partire dalle pochissime copie di DNA presenti in un campione. La reazione PCR quindi permette ai ricercatori (e agli investigatori) di aumentare esponenzialmente il numero di molecole di DNA eventualmente presenti in un campione; tale quantità può essere quindi analizzata, verificando se la sequenza dei nucleotidi nel campione è compatibile con il profilo del DNA di un soggetto. Teoricamente ad ogni ciclo di reazione PCR il numero di molecole di DNA dovrebbe raddoppiare; ciò si può esprimere come una funzione esponenziale, dove il tasso di amplificazione massimo desiderato è 2; tale valore viene chiamato efficienza della reazione.

2. L'ESPERIENZA DI URBINO: IL SISTEMA CYO

Normalmente si parte dall'assunto che a ogni ciclo di replicazione l'efficienza del sistema sia costante e massima, ma non è sempre vero; se nel campione ci sono impurità come tannini, flavonoidi o immunoglobuline l'amplificazione può essere inibita. Le impurità, chiamate inibitori, sono molto diffuse nell'ambiente, e incidono maggiormente nei campioni contenenti una minima quantità di acidi nucleici. Con una riduzione dell'efficienza di appena il 4 per cento (quindi con un'efficienza di 1,96 piuttosto che 2), per esempio, dopo 24-25 cicli di replicazione del DNA l'errore finale nel giudizio del-

la quantità di DNA effettivamente presente nel campione può essere del 400 per cento. Da qui la necessità di dotarsi di un sistema più affidabile, rispetto ai metodi standard, utilizzati ad oggi in tutti i laboratori di biologia molecolare. Il sistema messo a punto a Urbino, chiamato Cy0 (Guescini et al.), è di fatto un algoritmo matematico che tiene conto di diversi parametri cinetici e riesce a ridurre le fonti di errore che si hanno quando il materiale di partenza è poco e “inquinato”, come spesso accade nel caso di reperti della scena di un crimine. Valutare un campione di DNA con questo nuovo metodo rende la quantificazione più precisa e affidabile. Il metodo è liberamente disponibile a chiunque voglia utilizzarlo alla pagina web: <http://www.cy0method.org/index.php>. Tale lavoro ha riscontrato un notevole interesse internazionale, con circa 9000 download e 26000 accessi al sito sopra riportato.



Esempio di curva di amplificazione in real-time PCR; in ascissa è riportato il numero di cicli di amplificazione, mentre in ordinata la fluorescenza della quantità di DNA amplificato (in unità arbitrarie). F_0 è la quantità di DNA iniziale, che deve essere calcolata a partire dalla curva di amplificazione.

Le potenziali fonti di DNA possono essere le più svariate; si consideri che anche con il semplice contatto qualche cellula della nostra pelle si desquama, portando con sé il proprio DNA; da tale DNA è possibile, tramite PCR, ottenere un numero di acidi nucleici sufficiente per poter leggere la sequenza delle basi. Si è peraltro scoperto che il 99,9% del DNA è identico in tutti gli individui e soltanto il restante 0,1% di sequenza è responsabile delle caratteristiche che rendono proprio ciascun individuo e che, di fatto, ne consentono l'identificazione. In pratica vengono analizzate le VNTR (*Variable Num-*

ber Tandem Repeat) o le STRs (*Short Tandem Repeats*), ovvero corte ripetizioni in tandem di determinate sequenze di DNA. Sono brevi sequenze di DNA (da 6 a 40 nucleotidi) ripetute in tandem numerose volte (centinaia o anche migliaia di volte) in più loci cromosomici. Il numero di ripetizioni è altamente variabile (e quindi fortemente specifico) nei diversi individui. Numerose STRs e VNTR sono state validate e se ne è stabilito nelle diverse popolazioni umane la frequenza delle singole ripetizioni. Generalmente l'analisi da 9 a 15 STRs consente di raggiungere una percentuale di attribuzione con errore pari circa a $1/10^{12}$.

Senza entrare nei formalismo dei calcoli, vediamo come è possibile ottenere tale ordine di grandezza di probabilità di errore. La probabilità di avere una specifica VNTR dipende dal proprio specifico patrimonio genetico; ovviamente per ogni singolo sito contenente VNTR ci possono essere più varianti, che in questo caso sono il numero delle sequenze ripetute. Si consideri inoltre che se due eventi sono indipendenti, la probabilità che entrambi si verificano è data dal prodotto della probabilità dei singoli eventi; per esempio, nel lancio di una moneta, la probabilità che esca 'testa' al primo lancio è 0.5; al secondo è egualmente 0.5; quindi la probabilità che lanciando due monete si abbiano due teste è $0.5 \times 0.5 = 0.25$, cioè un caso su quattro. Se si considera la probabilità di una singola variante di VNTR pari a 0.16 (una stima prudenziale all'eccesso, in quanto si presuppone che per ogni VNTR si siano solo 6 possibili alternative), analizzando 15 VNTR si ottiene una probabilità di errore pari a 0.2^{15} ; tale numero è pari a circa $1/10^{12}$. Se per ciascuno VNTR ci sono più possibili alternative la probabilità diminuisce ulteriormente; se invece alcuni specifici VNTR si associano ad altri specifici VNTR posti in loci diversi, la probabilità aumenta.

È interessante notare come alcune ricerche mostrino come alcuni marcatori genetici siano associati all'etnia o addirittura alle caratteristiche somatiche del soggetto da cui proviene il DNA. Ciò significa che è possibile compiere una errata attribuzione del DNA di un reperto solo in un caso su mille miliardi di reperti analizzati; ossia, visto che sulla terra ora vivono, con una stima all'eccesso, 10 miliardi di persone, ci si attende un solo errore se si analizzano tutte gli uomini della terra, per cento volte di seguito! È chiaro che un simile livello di specificità è inarrivabile, utilizzando le classiche tecniche di analisi dei reperti. Ciononostante, è proprio l'estrema sensibilità il tallone d'Achille dei metodi biomolecolari; se sulla scena del delitto sono passate altre persone, se la vittima ha avuto contatti con qualcuno prima della morte (ed escludendo gli eremiti, tutti abbiamo contatti sociali...), se gli oggetti sono stati toccati da altri in passato, si ritrovano molteplici diversi pattern di DNA, i quali vanno necessariamente contestualizzati, al fine di non compiere gravi errori di interpretazione ed attribuzione.

Inoltre molto spesso si risale al DNA del presunto omicida, ma non si riesce a trovare la persona compatibile con quello specifico DNA. Questo è, ad esempio il caso di alcuni casi di cronaca nera italiani. Il tutto si complica in Italia in quanto, diversamente dalla maggior parte degli stati europei, non è stato ancora istituito un database dove sono depositate le *fingerprints* genetiche dei pregiudicati. Un esempio eclatante dei problemi che presenta l'elevatissimo pericolo di inquinamento del reperto lo si è avuto in Germania nel 2009. Da due anni la polizia tedesca rinveniva sulla scena di gravi delitti, tra cui anche l'omicidio di un poliziotto, lo stesso DNA di una persona di sesso femminile; alla fine si aveva il quadro di una specie di serial killer femminile detto "Phantom vom Heilbronn", che girava per la Germania commettendo delitti gravi. Già le erano stati attribuiti 40 casi con la conseguenza che non era stata svolta alcuna altra indagine per accertarne gli autori, perché la prova del DNA veniva considerata sufficiente. Poi per caso si è fatta la penosa scoperta che i bastoncini di ovatta usati per il prelievo del DNA erano forniti da un'unica ditta ed erano stati tutti contaminati del DNA di una operaia della ditta stessa! È quindi evidente che le analisi genetiche non sono di per sé risolutive, ma possono, se ben utilizzate, essere molto utili nel corso delle indagini. Per terminare, è sorprendente come Sherlock Holmes abbia, senza volerlo, posto le basi dell'analisi del DNA affermando che: "It has long been an axiom of mine that the little things are infinitely the most important" (Doyle 1981: 194).

BIBLIOGRAFIA

- Doyle, A.C. (1892), "A Case Of Identity", in *The Penguin Complete Sherlock Holmes*, London, 1981: 190-201.
- Guescini, M., D. Sisti, M.B. Rocchi, L. Stocchi, V. Stocchi (2008), "A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition", *BMC bioinformatics*, 9: 326.
- [Guescini, M., D. Sisti, R. Panebianco], "Cy0: A new approach in PCR quantification", <http://www.cy0method.org/index.php>.
- Mullis, K. (1998), *Dancing Naked in the Mind Field*, trad. it. *Ballando nudi nel campo della mente*, Milano, Baldini & Castoldi, 2000.
- Piattelli Palmarini, M. (2004), "«Così creai la fotocopiatrice del Dna» – Intervista allo scienziato Usa Kary Mullis", *Corriere della Sera*, 6 ottobre 2004, 19.

ABSTRACT

DNA is a long molecule, which is the basis of the chemistry of living beings. The DNA of an individual is the cause of its uniqueness, whether it is a unicellular organism, a plant or an animal. Until about two decades ago the study of DNA was very difficult, but now PCR (polymerase chain reaction) generates in a short time specific sequences of DNA which initially are present only in a small quantity. The possible sources of DNA can be wide-ranging; consider that in simple contact, cells of our skin may separate in flakes containing DNA; from such samples PCR can read sequences of the DNA. Generally DNA analysis allows us to reach a percentage of attribution with an error of about $1/10^{12}$ (which means an error every 1000 billion analyses). Nevertheless, the extreme ease of leaving traces of DNA is the Achilles heel of biomolecular methods: if at a crime scene other people have passed, if the victim had contact with someone else before his death, if the objects have been touched by others in the past, different patterns of DNA will be found, which must, of course, be contextualized in order not to make serious errors of interpretation.