

Il sequenziamento parallelo, game changer nello studio degli acidi nucleici

A.Valentini

Università della Tuscia, Viterbo

Poiché gli acidi nucleici contengono pressoché tutte le informazioni di una cellula o di un organismo in ordine sequenziale, la conoscenza di tale ordine da luogo alla possibilità di comprendere, tra l'altro, le caratteristiche specifiche di tale cellula o organismo distinguendolo da altri. L'utilizzazione delle tecniche di sequenziamento per la diagnostica si è purtroppo sviluppata lentamente all'inizio per tre fattori limitanti: la necessità di utilizzare primers specifici di una specie (e quindi biased nei confronti delle altre specie non target); la possibilità di analizzare solo un breve tratto del genoma e quindi con una scarsa informatività, la scarsità di informazioni presenti nei database di riferimento. Intorno al 2005 sono state messe in commercio macchine di sequenziamento che consentivano il processamento in parallelo di migliaia, e poi di milioni di tratti di acidi nucleici contemporaneamente. Da quel momento il numero di basi sequenziate ha iniziato a crescere in modo esponenziale, e i costi a calare in modo altrettanto vertiginoso. Tuttavia questa mole di dati prodotti ha creato ulteriori problemi: la crescita dei dati è superiore alla crescita della velocità di processamento dei computer (superiore alla legge di Moore). In questi anni si sviluppa in modo particolare una giovane scienza, la bioinformatica, che diventa fondamentale per poter sviluppare ed applicare algoritmi capaci di utilizzare la enorme massa di dati a disposizione. Infine vengono "prestati" alla bioinformatica anche speciali sistemi di processo informatico in parallelo originariamente sviluppati per il mercato dei videogiochi, le Graphic Processing Units (GPU) che ad un costo contenuto sono capaci di processare centinaia di migliaia di programmi in parallelo. Queste conquiste concettuali e tecnologiche stanno cambiando radicalmente il settore della diagnostica microbiologica. Il primo passo è la generazione di database controllati delle sequenze genomiche. Sono presenti diverse iniziative di consorzi internazionali, sia per la creazione di database di genomi completi che di regioni ristrette ma particolarmente adatte per la discriminazione di specie o genere. Il confronto con i genomi completi è sicuramente il sistema di diagnostica più completo, ma in pratica presenta diverse criticità: i database relativi sono ancora carenti, il processamento dei dati è complesso (si devono ricercare alcuni milioni di sequenze su database di diversi miliardi di basi), se il campione analizzato presenta diverse specie si può avere una scarsa copertura delle specie meno abbondanti. Se il sequenziamento di zone diagnostiche risolve alcuni dei problemi suesposti, a sua volta ne crea altri: i primers utilizzati nella pre-amplificazione non sono del tutto "universali" e quindi si rischia di perdere alcune specie, i database pubblici già preparati dai consorzi ("chiavi in mano") sono spesso vecchi e quindi è necessario avere alcune competenze di bioinformatica per crearne di propri aggiornati. In conclusione il sequenziamento parallelo offre possibilità diagnostiche e conoscitive impensabili fino a pochi anni fa, ma necessita di strumenti sofisticati, non tanto per l'esecuzione delle analisi di laboratorio che sono abbastanza automatizzate o affidabili in outsourcing, ma soprattutto per l'analisi bioinformatica che richiede macchine e competenze particolari non sempre facilmente reperibili.