



UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI BIOTECNOLOGIE

Strada Le Grazie, 15 – Ca' Vignal 1 – 37134 Verona (Italy)

Massimo Delledonne

Il recentissimo sviluppo delle “next generation sequencing technologies” ovvero delle tecnologie di sequenziamento massivo parallelo del DNA, sta radicalmente cambiando lo scenario del sequenziamento dei genomi. Se il sequenziamento del genoma umano, ottenuto con le tecnologie tradizionali (Sanger), è costato diversi anni di lavoro e 10 miliardi di dollari, il risequenziamento di un genoma umano con le tecnologie oggi disponibili richiede poche settimane e costa ormai poche decine di migliaia di dollari. Si parla di “risequenziamento” perché queste nuove tecnologie, estremamente potenti, hanno la caratteristica di produrre sequenze molto corte, sequenze che difficilmente riescono ad essere assemblate da sole per ricostruire l'intero genoma. Tuttavia, queste brevi sequenze possono con relativa facilità essere allineate su un genoma di riferimento, ottenuto con tecnologie tradizionali, e confrontate per determinare le differenze strutturali e funzionali fra il genoma di riferimento e il genoma oggetto dello studio.

Grazie alla disponibilità di un genoma di riferimento per la vite, il clone Pinot nero 40024 (PN40024) decodificato nel 2007 dal consorzio italo-francese Vigna-Vigne di cui fa parte anche l'Università degli Studi di Verona, e di un sequenziatore di ultima generazione *Illumina Genome Analyzer IIx*, acquisito in comproprietà con l'Istituto di Genomica Applicata di Udine, l'Università degli Studi di Verona ha potuto completare il primo risequenziamento di una varietà locale di vite.

Sequenziamento del genoma

Nella prima fase del progetto, il genoma del clone 48 di Corvina è stato analizzato dall'Università degli Studi di Verona in collaborazione con l'Istituto di Genomica Applicata di Udine. Utilizzando il sequenziatore *Illumina Genome Analyzer IIx* sono state generati 230 milioni di sequenze corte (“reads” da 36 basi in modalità “paired-ends”), per un totale di oltre 8.4 gigabasi (Gb) di dati. Dato che il genoma di *Vitis vinifera* è circa 450 milioni di basi, è stata ottenuta una copertura del genoma pari a circa 19 volte (19X).

I dati di sequenza sono stati allineati al genoma di riferimento PN40024. La mappatura delle *reads* sul genoma di riferimento ha permesso l'individuazione di 329,000 polimorfismi del DNA (single nucleotide polymorphisms, SNP) che interessano quasi 14,000 geni e che possono potenzialmente modificarne l'attività. Come atteso, è stato confermato che il genoma di Corvina non mostra segni di inincrocio tipici delle specie sottoposte a miglioramento genetico intensivo: Il 75% degli SNP rilevati si trova infatti in eterozigosi e conferma che le due copie del genoma presenti in ogni cellula differiscono in molti punti.

La mappatura delle *reads* ha inoltre permesso di identificare quasi 500 delezioni, di lunghezza media di circa 46 kilobasi, e 20 duplicazioni, di lunghezza media di 43 kilobasi. La differenza fra il genoma di Corvina e il genoma di PN40024 è stata stimata nell'ordine delle 25 Mb (milioni di basi).



UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI BIOTECNOLOGIE

Strada Le Grazie, 15 – Ca' Vignal 1 – 37134 Verona (Italy)

Sequenziamento del trascrittoma

La sequenza del genoma di Corvina fornisce importanti informazioni dal punto di vista evolutivo e strutturale, informazioni che sono di grande interesse per la ricerca di base ma che hanno un limitato valore applicativo in assenza di una precisa mappa dei geni in essa contenuti. Risulta pertanto fondamentale decifrare la sequenza genomica e produrre una solida annotazione dei geni di Corvina e della loro struttura. Dato che la sequenza dell'RNA rispecchia la sequenza del DNA da cui è stata trascritta, l'analisi dell'intera collezione di RNA (trascrittoma) permette ai ricercatori di determinare come e quando ciascun gene è impiegato oppure no, e la sua precisa struttura. Negli organismi multicellulari infatti, ciascuna cellula contiene lo stesso genoma, e quindi lo stesso set di geni, ma cellule diverse manifestano profili diversi nell'espressione dei geni. Queste differenze sono responsabili delle varie proprietà e comportamenti che cellule di diversi tessuti mostrano. Per esempio, le cellule di una foglia attivano il set di geni necessario a fissare l'energia del sole mediante la fotosintesi clorofilliana, mentre le cellule della bacca attivano il set di geni coinvolto nel suo sviluppo e nella sua maturazione. Per questo, collezionare e comparare trascrittomi di diversi tipi di cellule, è fondamentale per comprendere i geni che definiscono uno specifico tipo di cellula, come questa cellula funziona, e come variazioni ambientali influenzano il normale livello di attività dei geni. Inoltre, allineando il trascrittoma di ciascuna cellula sul genoma, è finalmente possibile ricostruire la struttura di tutti i geni di quell'organismo

Il Centro di Genomica Funzionale Vegetale dell'Università degli Studi di Verona, specializzato nello studio dell'espressione dei geni, ha prodotto 59.2 milioni di sequenze dal trascrittoma di bacche di Corvina prelevate durante diverse fasi di maturazione, generando oltre 2.2 gigabasi (Gb, o miliardi di basi) di sequenza, e le ha poi allineate con successo sul genoma di PN40024.

Circa 48.5 milioni di sequenze sono state mappate in corrispondenza dei geni conosciuti, mentre 1.5 milioni di sequenze sono state mappate in regioni non associate a modelli genici, e che non si sapeva fossero trascritte. Grazie all'ausilio di computer estremamente potenti, il gruppo di ricerca del Centro ha sviluppato un software che ha permesso di assemblare queste sequenze in 479 modelli genici fino ad oggi sconosciuti.

A questi nuovi geni, i ricercatori hanno affiancato numerose varianti proteiche generate da eventi di *splicing* alternativo osservati durante la maturazione dell'uva. Negli eucarioti infatti, la produzione di proteine segue un meccanismo simile alle operazioni di taglio e cucito. Dopo essere stato trascritto dal DNA, ogni RNA messaggero subisce un processo di "maturazione" (*splicing*) costituito dalla rimozione di alcune porzioni, gli introni e la giunzione delle porzioni rimanenti, gli esoni. Da diversi anni è noto che, soprattutto nel regno animale, oltre allo *splicing* canonico (che consiste nella eliminazione di tutti gli introni ed il mantenimento di tutti gli esoni), alcuni geni possono subire *splicing* alternativi che portano alla produzione di RNA messaggeri maturi diversi a partire da un unico precursore. Pertanto, a partire da un unico gene è possibile sintetizzare più proteine che possono differire per funzione, localizzazione o altre proprietà, semplicemente variando il numero di esoni che vengono mantenuti nell'RNA maturo. Fra le oltre 41.500 giunzioni di *splicing* rilevate in Corvina, circa 450 corrispondono a eventi di *splicing* alternativo in 345 geni coinvolti nella maturazione della bacca.

Altra importante scoperta è stata l'individuazione di un piccola inserzione in quello che si pensava fosse uno pseudogene. Sebbene mantengano a volte la struttura tipica dei geni (promotore, siti di *splicing*), queste sequenze ancestrali non sono in grado di generare un prodotto proteico funzionale, spesso a causa di mutazioni genetiche consolidate durante l'evoluzione. Ebbene, la



UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI BIOTECNOLOGIE

Strada Le Grazie, 15 – Ca' Vignal 1 – 37134 Verona (Italy)

mutazione identificata dimostra che mentre in PN40024 questo gene è inattivo (è appunto un pseudo gene), in Corvina questo gene è funzionale: esso è espresso, e codifica per un fattore di trascrizione bHLH, direttamente coinvolto nella sintesi dei flavonoidi e di altre molecole legate alle caratteristiche tipiche dell'uva Corvina.

Infine, i ricercatori veronesi si sono concentrati su quelle sequenze ottenute dal trascrittoma di Corvina che non avevano corrispondenza con il genoma di PN40024 e che inizialmente, proprio per questo, erano state scartate. Seguendo uno schema di lavoro molto simile a quello utilizzato per assemblare il genoma del Panda (pubblicato il 21 gennaio 2010 sulla rivista Nature), i ricercatori hanno implementato un software che ha permesso loro di assemblare quelle sequenze in "contigs" corrispondenti a 187 geni, che risultano espressi in Corvina e invece del tutto mancanti dal genoma di PN40024. Per 33 di questi geni è già stata dimostrata la funzionalità.

In sintesi, il sequenziamento del trascrittoma di Corvina e la sua comparazione con il genoma di riferimento hanno permesso di scoprire oltre 500 nuovi geni, molti dei quali mancano completamente nel genoma di PN40024, e di identificare ben 450 varianti trascrizionali in 345 geni coinvolti nello sviluppo della bacca. Questi dati confermano una peculiarità e complessità trascrizionale del processo di maturazione dell'uva Corvina davvero formidabile

Genoma e Trascrittoma: cosa sono

Il genoma è composto da acido deossiribonucleico, più noto con la sigla di DNA, e contiene le istruzioni necessarie alla cellula per organizzarsi e sopravvivere. Le istruzioni contenute in questo codice devono però essere trascritte in molecole corrispondenti di acido ribonucleico, meglio conosciuto come RNA, a cui ci si riferisce come trascritti. **Il trascrittoma** è l'insieme di tutti i trascritti presenti in una data cellula. Esistono diversi generi di RNA. Il più presente, chiamato **RNA messaggero** o mRNA, svolge un ruolo fondamentale nella sintesi delle proteine. L'RNA messaggero viene trascritto a partire dai geni, che racchiudono le informazioni relative alle proteine cellulari, e trasportato fino ai ribosomi. Questi sono macchine molecolari capaci di leggere la sequenza di lettere chimiche dell'mRNA e "tradurla" in una sequenza corrispondente di aminoacidi, gli elementi fondamentali che formano le proteine. Ogni mRNA viene trascritto da un gene e assemblato in una specifica proteina.

Verona, 26 febbraio 2010