



# I PRONIPOTI DI MENDEL NELL'ERA POST-GENOMICA

In pochi giorni si può conoscere il profilo di espressione di una pianta ma lo studio globale di proteine e metaboliti è ancora ai primi passi

PIERDOMENICO PERATA, AMEDEO ALPI

**G**LI ESPERIMENTI EFFETTUATI DAL MONACO Gregor Johann Mendel sull'ereditarietà dei caratteri sono noti a tutti. L'incrocio tra piante di pisello a semi lisci e piante a semi rugosi consente, osservando l'aspetto esteriore (fenotipo) dei semi ibridi e della loro successiva progenie, di delineare un'importante legge della genetica moderna. Ciò che Mendel non avrebbe potuto studiare, invece, è la natura della mutazione che causava il fenotipo rugoso. Solo nel 1990, infatti, è stata svelata l'origine di questa mutazione, che verosimilmente è causata da un trasposone. I trasposoni sono tratti di DNA in grado di spostarsi lungo un cromosoma o addirittura da un cromosoma all'altro, modificandone la sequenza. Fu Barbara McClintock a descriverli, studiando il mais e le sue peculiari mutazioni che causano la colorazione delle pannocchie. Per questa scoperta la McClintock ha ricevuto il premio Nobel nel 1983.

Le possibilità di approfondimento di questi fenomeni genetici incontrano un ostacolo nell'impossibilità di studiarli, a livello globale, utilizzando i sistemi vegetali che ne hanno consentito la scoperta. Infatti se si desidera non solo valutare le conseguenze di una mutazione sul fenotipo di una pianta, ma anche analizzare gli effetti di tale mutazione sull'attività del network di regolazione genica, occorre impiegare specie vegetali i cui genomi siano stati ampiamente sequenziati, per consentire lo sviluppo di strumenti di analisi post-genomica. Il nuovo millennio vede infatti una specie vegetale come principale protagonista: *Arabidopsis thaliana*. Questa piccola pianta alta pochi centimetri, con soli cinque cromosomi, rappresenta il sistema modello per migliaia di ricercatori in tutto il mondo, che sfruttano la semplicità con la quale è possibile ottenere mutanti, incroci e piante transgeniche. *Arabidopsis thaliana* è una piccola pianta erbacea che appartiene alla famiglia delle brassicacee. È tipica delle regioni temperate dell'emisfero boreale e cresce spontanea su terreni incolti e aridi. Al genere *Arabidopsis* appartengono diverse specie, ma *Arabidopsis thaliana* – che fu scoperta nel sedicesimo secolo in Germania sulle montagne di Harz da Johannes Thal, da cui deriva il nome della specie – è la più conosciuta e la più impiegata ai fini della ricerca. Il suo genoma è organizzato in cinque cromosomi e risulta piccolo se confrontato con quello delle maggiori specie di interesse agronomico quali riso, pomodoro, mais e frumento, i cui genomi sono rispettivamente tre, sette, venti e centoventi volte più grandi<sup>1</sup>. Le regioni genomiche sequenziate contengono 25.498 geni. Le funzioni di circa il 69% dei geni sono state classificate in base a similarità di sequenza con proteine a funzione nota di altri organismi (solo il 9% è stato caratterizzato sperimentalmente). Il restante 30% dei predetti 25.498 pro-



Polline di Arabidopsis fotografato al microscopio elettronico.

JÜRGEN BERGER / HIKO SCHOOF / ELECTRON MICROSCOPY UNIT, MAX PLANCK INSTITUT



CORTESIA AMERICAN PHILOSOPHICAL SOCIETY LIBRARY, BARBARA MCCLINTOCK PAPERS

Aprile 1963, Barbara McClintock nel suo studio mentre lavora sul mais.

dotti genici, che comprendono sia proteine vegetali specifiche sia proteine con similarità a geni di altri organismi la cui funzione non è ancora nota, non è stato assegnato ad alcuna categoria funzionale. Contrariamente ad altri membri della stessa famiglia, come ravanello, cavolfiore, verza e rapa, *Arabidopsis* non riveste alcun interesse agronomico, ma rappresenta la pianta modello per eccellenza per gli studi di fisiologia e genetica vegetale. Si sviluppa e risponde a stress e malattie in maniera molto simile alla maggior parte delle piante coltivate. Ha un ciclo vitale, da seme a seme, molto breve. È piccola, pertanto adatta a crescere in spazi ristretti come quelli di un laboratorio, e una singola pianta è in grado di produrre fino a 10.000 semi. Ha dimensioni molto ridotte (0,5 cm di altezza) e per questo si presta bene a esperimenti genetici su ampia scala. Non sorprende quindi che, in pochi anni, si sia formata una comunità scientifica internazionale impegnata nello studio di questa pianta, che ha impresso una svolta significativa nel campo della ricerca vegetale. Tradizionalmente le conoscenze sulla morfologia e sulla fisiologia delle piante derivavano da studi effettuati su specie interessanti dal punto di vista agronomico. Sebbene in questo modo siano state raccolte numerose informazioni, questo approccio portava inevitabilmente a una dispersione delle energie che potevano essere invece convogliate in maniera più proficua concentrando gli sforzi sullo studio di un'unica pianta. Diverse specie sono state proposte come possibili candidate – ad esempio mais, pomodoro, pisello, riso, orzo, petunia, bocca di leone

– ma non si riusciva a trovare un accordo su quale fosse la più rappresentativa per studiare processi comuni a tutte le piante.

L'attenzione della comunità scientifica si è focalizzata su *Arabidopsis* verso la fine degli anni Settanta. L'interesse è cresciuto enormemente all'inizio degli anni Ottanta con la pubblicazione di numerosi articoli che ne sottolineavano il potenziale valore per la ricerca, con la determinazione della prima mappa dettagliata dei suoi cinque cromosomi, con la dimostrazione che aveva un genoma piccolo e quindi estremamente adatto per analisi genetiche e con la realizzazione di protocolli di produzione di esemplari transgenici. Negli anni Novanta, quindi, *Arabidopsis* è divenuta la pianta modello per gli studi di fisiologia, biochimica e genetica vegetale. Nello stesso periodo sono stati delineati diversi importanti obiettivi da raggiungere attraverso la collaborazione internazionale, quali la saturazione del genoma con mutazioni inserzionali e il sequenziamento completo entro la fine del decennio. Con l'attivazione dell'*Arabidopsis Genome Initiative* – che ha avuto inizio nel 1996 e ha coinvolto 15 laboratori tra Europa, Stati Uniti e Giappone – si è arrivati alla pubblicazione nel 1999 della sequenza dei cromosomi 2 e 4 per finire nel dicembre 2000 con la sequenza completa dei cromosomi 1, 3 e 5<sup>1</sup>. *Arabidopsis thaliana* rappresenta il primo organismo vegetale del quale è stato completato il sequenziamento del genoma, poi più recentemente è stata sequenziata un'importante specie coltivata, il riso<sup>2,3</sup>.



CORTESIA AMERICAN PHILOSOPHICAL SOCIETY LIBRARY. BARBARA MCCLINTOCK PAPERS

Una foto di Cold Spring Harbor nell'inverno del 1942 realizzata da Barbara McClintock

### I miracoli del trascrittoma

Il sequenziamento del genoma di *Arabidopsis* ha colmato in parte il divario che separa la biologia animale da quella vegetale. La conoscenza della sequenza ha infatti consentito lo sviluppo di strumenti per la post-genomica vegetale. Come è noto, la post-genomica studia gli eventi che sono diretta conseguenza delle sequenze genomiche, e quindi in primis l'insieme degli RNA messaggeri presenti in un sistema biologico in un particolare stato fisiologico (trascrittoma), il conseguente insieme di proteine (proteoma) e infine l'insieme dei metaboliti risultanti dall'attività metabolica svolta dalle proteine nell'ambito cellulare (metaboloma). Anche se è divenuto consuetudine considerare l'era della post-genomica come caratterizzata dalla triade trascrittoma-proteoma-metaboloma, occorre essere realisticamente consapevoli che solo la trascrittomica ha raggiunto una maturità tecnologica tale da poterla considerare una scienza consolidata. Se prima dell'avvento delle tecnologie trascrittomiche il tempo necessario per studiare l'espressione di appena una mezza dozzina di geni era di circa una settimana, oggi in tale arco di tempo è possibile, anche per un piccolo laboratorio, ottenere dati relativi all'espressione di oltre 25.000 geni. Il costo di queste tecnologie rappresenta una limitazione al loro impiego, ma la spesa è elevata solo in apparenza e diventa del tutto ragionevole se ripartita sul numero di dati ottenuti. La principale tecnologia impiegata per lo studio del trascrittoma delle piante è identica a

quella applicata ad altri sistemi viventi, uomo incluso: i microarray.

I microarray utilizzano centinaia o migliaia di sonde di DNA disposte su una superficie solida di dimensioni ridotte, spesso un quadrato di meno di un centimetro di lato. Questo sistema consente di valutare la presenza e l'abbondanza di specifici trascritti in un estratto biologico. Le sonde di DNA presenti in un microarray possono essere cDNA o oligonucleotidi con sequenze complementari alla sequenza bersaglio (la sequenza del trascritto ricercato). Questi oligonucleotidi sono sintetizzati direttamente sulla superficie del microarray o sono depositati e immobilizzati in posizioni specifiche su un supporto solido, a formare una griglia ordinata. I campioni da analizzare, cDNA o cRNA derivati dall'RNA messaggero estratto dal campione, sono resi fluorescenti o radioattivi o marcati con biotina prima dell'ibridazione sul supporto solido. Quando i campioni vengono applicati al microarray, quindi, l'ibridazione avviene tra la sonda adesa al supporto e il bersaglio mediante riconoscimento tra sequenze complementari. L'intensità del segnale fluorescente o radioattivo riflette l'abbondanza delle molecole bersaglio e la compatibilità di legame fra la sonda e le molecole bersaglio stesse. L'impiego di microarray è divenuto il metodo predominante per l'analisi contemporanea dell'espressione genica in vari processi biologici delle piante, come il mantenimento dei ritmi circadiani, lo sviluppo di resistenza a malattie, le risposte a stress ambientali, lo sviluppo dei frutti e del seme, la fotomorfogenesi e

molti altri. Inoltre i microarray non sono stati sviluppati solo per *Arabidopsis*, ma anche per specie quali la fragola, il riso, l'orzo, il pioppo e il pino. Una tipologia particolare di microarray è rappresentato dai GeneChip, che si basano sulla sintesi *in silico*, direttamente sulla superficie del microarray, degli oligonucleotidi che fungono da sonda. Questa tecnologia è di fatto uno standard per quanto concerne i microarray di specie modello. Rispetto al primo GeneChip di *Arabidopsis*, progettato nel 2000, il numero di sequenze sonda su unità di superficie di un singolo GeneChip è raddoppiato. L'ATH1, un GeneChip *genome-wide* sviluppato recentemente, contiene sequenze in grado di coprire il 90% del genoma di *Arabidopsis* <sup>4</sup>.

Anche se l'analisi del trascrittoma tramite i microarray è ancora nelle sue fasi iniziali, sono già evidenti le potenzialità di questo approccio, che è stato impiegato per caratterizzare l'espressione genica regolata da ormoni o altri trattamenti, per determinare gli elementi regolatori di geni, per caratterizzare le funzioni geniche. L'analisi dell'espressione genica in mutanti genetici non soltanto caratterizza i fenotipi molecolari dei mutanti stessi al livello quantitativo, ma fornisce anche una grande quantità di informazioni che può essere usata per definire la rete di trasmissione dei segnali (*signaling*) conseguente alla mutazione in esame.

Le analisi di espressione genica che spaziano sull'intero genoma offrono opportunità uniche per studiare le interazioni tra geni e vie metaboliche e caratterizzare le reti regolatrici dei diversi geni. Infatti, anche se tradizionalmente i processi di trasmissione dei segnali cellulari sono stati descritti come una serie di eventi lineari, è ormai evidente che questi processi possono consistere di una rete di eventi. Nel loro ambiente naturale, le piante sono sottoposte a una varietà di stress abiotici e biotici – quali vento, pioggia, il contatto meccanico, siccità, stress termici, stress da insetti, ferite inflitte dai fitofagi e infezione da agenti patogeni – e recentemente è stata dimostrata l'esistenza di interazioni (*cross-talk*)

fra le vie che sono coinvolte nelle risposte ad agenti patogeni e a vari stress abiotici.

Gli esperimenti di microarray sono stati usati per fare luce sui meccanismi della regolazione trascrizionale. L'analisi di gruppi di geni che presentano profili

di espressione molto simili consente il loro raggruppamento in cluster. Questi profili possono indicare la regolazione coordinata di una serie di geni, perciò questo metodo è stato impiegato per identificare i geni co-regolati e, congiuntamente a conoscenze acquisite in era pre-genomica, può essere usato per raggruppare i geni in vie di trasduzione del segnale. Le analisi delle sequenze dei promotori di geni potenzialmente co-regolati può essere usata per identificare



Un particolare dell'infiorescenza di *Arabidopsis* fotografata al microscopio.

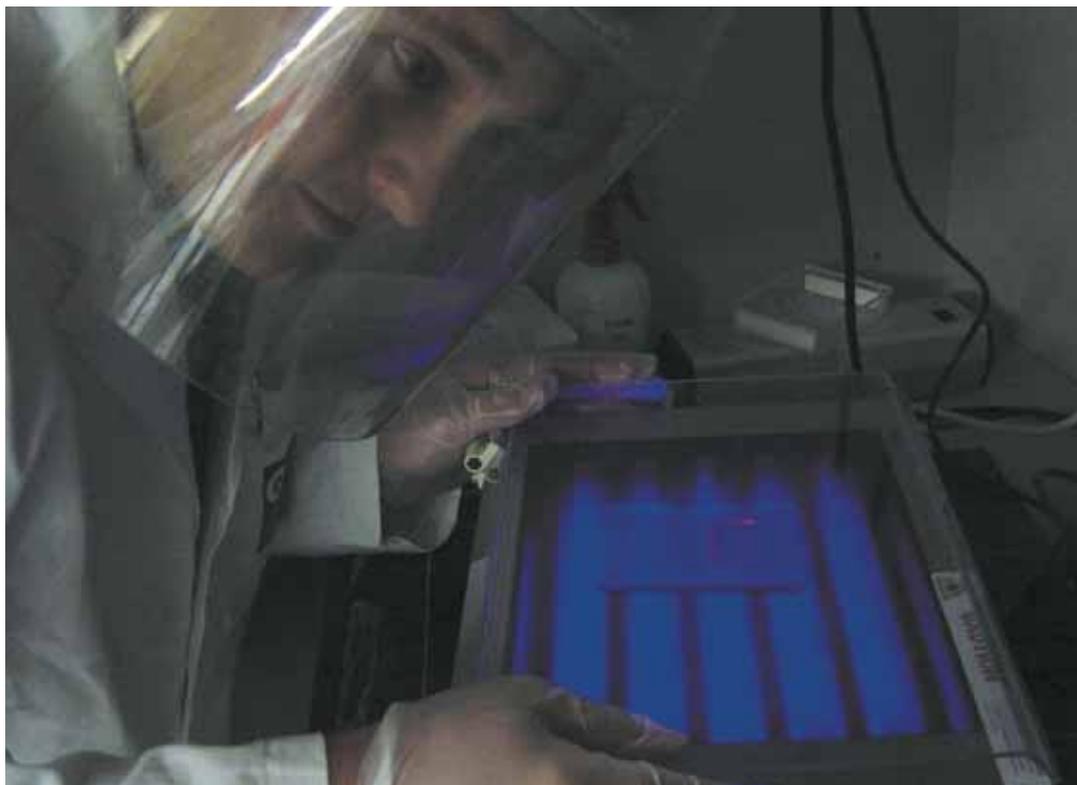
CORTESIA JURGEN BERGER, ELECTRON MICROSCOPY UNIT, MAX PLANCK INSTITUT

sequenze comuni. Il radicale cambiamento di approccio allo studio dell'espressione genica, dallo studio di singoli o pochi geni preventivamente selezionati allo studio del profilo di espressione genica a livello di intero genoma, fornisce una prospettiva di studio precedentemente inimmaginabile. Una mole di informazioni ancora più notevole può essere tratta dall'integrazione di studi di espressione genica, proteomica, metabolomica, statistica e bioinformatica.

### I limiti della proteomica

Il termine proteomica dovrebbe indicare, al pari della trascrittomica, una nuova scienza – o, meglio, una nuova metodologia scientifica – atta allo studio globale di tutte le proteine presenti in una cellula, tessuto o organo. In realtà, si osserva spesso un abuso della parola, che talvolta viene utilizzata impropriamente per indicare tecniche di separazione elettroforetica di proteine, ad esempio tramite la separazione bidimensionale in cui le proteine migrano prima separandosi in funzione del proprio punto isoelettrico, poi secondo il proprio peso molecolare.

Seppure è innegabile che l'elettroforesi bidimensionale sia parte della proteomica, la semplice descrizione di pattern elettroforetici è lungi dal poter essere considerata un approccio proteomico. La proteomica è infatti un esame sistematico dell'in-



CORTESIA KEIN MCCRAITH, UNIVERSITY OF QUEENSLAND

Esame con un illuminatore a luce ultravioletta di un gel con campioni di DNA vegetale.

sieme delle proteine presenti in un sistema biologico. Mentre la trascrittomica analizza molecole chimicamente omogenee (gli RNA messaggeri), nel caso della proteomica l'approccio sistematico è reso complesso dall'eterogeneità delle proteine, a partire dalla difficoltà di definire protocolli di estrazione in grado di garantire la rappresentatività del campione.

Molte proteine sono infatti legate a membrane, compartimentate, solubili a differenti pH e questo rende quasi impossibile l'estrazione contemporanea dell'intero proteoma. I protocolli per la separazione elettroforetica delle proteine e per la loro successiva identificazione sono a un buono stato di sviluppo e la biologia vegetale si avvantaggia dei progressi della proteomica umana. I limiti della proteomica, comunque, sono ben esemplificati dai numeri: le tecnologie attualmente disponibili consentono di separare circa 500-2.500 proteine di *Arabidopsis* in un singolo gel elettroforetico, a fronte di una stima largamente per difetto di 25.000 proteine presenti nella pianta.

Il progresso in questo campo richiede quindi un forte avanzamento nelle tecnologie per la separazione delle proteine, oltre a una maggiore disponibilità di informazioni sulle funzioni delle singole proteine<sup>5</sup>. La lacunosità dell'annotazione del genoma di *Arabidopsis* – che in buona parte codifica per proteine la cui fun-

zione non è nota – si ripercuote negativamente sulle possibilità di interpretazione dei risultati di trascrittomica e di proteomica, come verrà discusso nel paragrafo relativo alla bioinformatica.

### La sfida della metabolomica

Gli effetti dell'attività trascrizionale si riflettono sul proteoma e l'attività di quest'ultimo – che comprende, ad esempio, enzimi e trasportatori – risulta nella produzione di un largo numero di metaboliti, stimati nell'ordine delle centinaia di migliaia<sup>5</sup>. I metaboliti delle piante svolgono un numero rilevante di funzioni, da quelle relative al metabolismo primario (ad esempio intermedi del metabolismo respiratorio e fotosintetico) alla complessità del metabolismo secondario, responsabile di molte delle diversità e peculiarità delle singole specie vegetali. Il metabolismo secondario è infatti responsabile della produzione di metaboliti coinvolti nei processi di resistenza a batteri, funghi patogeni e insetti, nella risposta agli stress abiotici, nel colore, gusto e profumo di fiori e frutti. La natura e il livello di questi metaboliti sono certamente determinati dall'attività del genoma, ma anche dall'interazione di quest'ultimo con l'ambiente, che condiziona la trascrizione dei geni e quindi anche il proteoma e l'insieme risultante di metaboliti. Il numero di metaboliti presenti

nelle piante rende la sfida complessa ed è necessario ammettere che la metabolomica è ancora nelle fasi più precoci di sviluppo.

Esistono difficoltà metodologiche per isolare e quantificare le diverse classi di metaboliti, spesso da estrarre con procedure differenziate, per non menzionare le problematiche di quantificazione delle migliaia di composti volatili prodotti dalle piante. La principale difficoltà è rappresentata dal fatto che, per riportare un esempio reale, dei circa 5.000 metaboliti probabilmente presenti in una foglia di *Arabidopsis* solo il 10% è rappresentato da molecole note. Come sviluppare protocolli metabolomici (estrazione, purificazione, identificazione) per molecole la cui esistenza è solo predetta? Lo sviluppo della metabolomica è d'altra parte indispensabile per la corretta integrazione dei dati di trascrittomica, proteomica e, appunto, metabolomica. Solo un reale sviluppo delle «omiche» ancora in deficit tecnologico (proteomica e metabolomica) consentirà di esplorare le interazioni tra i diversi livelli di regolazione dei viventi.

### Il contributo della bioinformatica

Lo sviluppo delle «omiche» risulta nella produzione di una quantità di dati superiore alle possibilità di analisi. Il passaggio da sperimentazioni che riguardano l'espressione di poche decine di geni alle attuali panoramiche sul trascrittoma pone il ricercatore nella condizione di dover rivedere il proprio metodo scientifico. Infatti non è più accettabile la costruzione di ipotesi basate sull'espressione dei soli geni di interesse, essendo ormai gli esami del trascrittoma tramite microarray alla portata di molti laboratori.

La necessità di tener conto non solo delle variazioni più o meno attese dei geni di interesse, ma di inserire tra le conclusioni e le ipotesi derivanti dagli esperimenti anche l'espressione, magari eclatante, di altri geni apparentemente non correlati al fenomeno studiato, porta alla sempre maggiore esigenza di disporre di software in grado di filtrare i dati, consentendo una gestione più abbordabile del sistema sperimentale. I software in grado di elaborare i dati di espressione (ad esempio escludendo i geni poco espressi), di assegnare funzioni a gruppi di geni (ricostruire vie metaboliche sulla base dei dati di trascrittoma) e anche di predire le modalità di espressione di geni di interesse richiedono spesso la disponibilità di banche dati il più possibile aggiornate <sup>6,7</sup>.

La rapida obsolescenza delle informazioni affligge le «omiche» e ne condiziona lo sviluppo quando, ad esempio, geni annotati come coinvolti in una risposta fisiologica – magari per omologia di sequenza con altri organismi – risultano in realtà in grado di reagire anche ad altri stimoli e spesso il numero di condizioni a cui il gene reagisce è ben

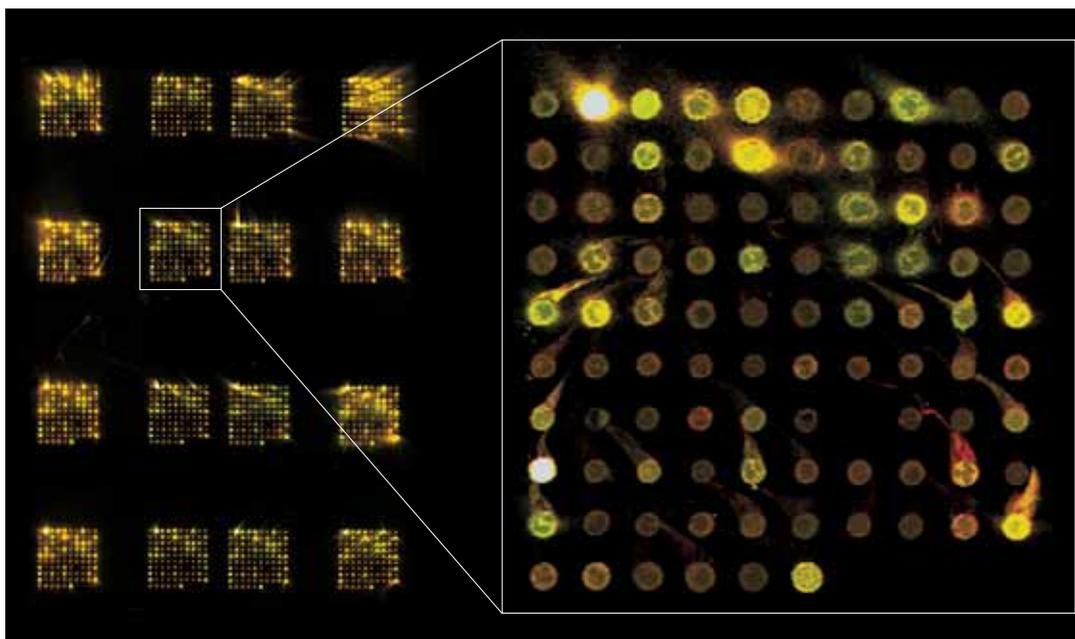
superiore alle attese. Se quindi il software assegna alla condizione «stress» (per omologia di sequenza o sulla base di evidenze pre-genomiche) la funzione di un gene particolarmente espresso, senza tener conto di ulteriori dati di trascrittomica che evidenziano come il gene sia spesso espresso in relazione a uno stato fisiologico differente dalla situazione di stress (ad esempio durante la fioritura della pianta), le conclusioni cui il software giungerà non saranno corrette. I dati di trascrittomica stanno, di fatto, riscrivendo in modo dinamico le annotazioni del genoma e diviene quindi un imperativo disporre di software in grado di interrogare banche di dati di espressione genica.

L'imponente mole di dati di trascrittoma che viene prodotta ogni giorno rischierebbe di essere parzialmente fine a se stessa se non venissero definiti dei criteri da seguire nella descrizione dello schema sperimentale impiegato per ottenere i dati. Infatti, mentre il genoma può essere considerato indipendente dall'ambiente – e quindi i dati sui genomi dipendono dall'individuo sequenziato, ma non dalle condizioni ambientali, mutevoli, in cui esso vive – i dati di trascrittoma sono per loro stessa natura strettamente dipendenti da fattori quali lo stadio di sviluppo della pianta, il tessuto in esame, le condizioni esterne in cui l'organismo vive nell'attimo in cui avviene il prelievo del materiale biologico per l'analisi.

Solo se i dati di trascrittoma sono accompagnati da una meticolosa descrizione di questi fattori, allora i risultati dell'analisi non saranno utili soltanto allo sperimentatore, che probabilmente ne utilizzerà una parte minima, ma a tutta la comunità scientifica che potrà interrogare i dati altrui per ottenere indicazioni sull'espressione di geni di proprio interesse nelle condizioni sperimentali definite altrove. Nuovi software specifici per *Arabidopsis* sono attualmente disponibili ed è facile prevedere che presto i ricercatori di biologia vegetale potranno contare su vaste banche di dati di espressione in grado non solo di definire come-dove-quando un gene si esprime, ma anche di muovere i primi passi nella complessità delle interazioni tra geni.

### Verso la pianta virtuale?

L'esistenza di software in grado di simulare la crescita e lo sviluppo di una pianta, anche in tre dimensioni, suggerisce la possibilità che queste simulazioni non siano condannate a basarsi su informazioni relativamente semplici, come la definizione dei parametri di crescita di un albero che sfugge all'ombra creata da un altro albero vicino <sup>8</sup>. Il traguardo ideale prevede la possibilità di invertire la prospettiva attuale che mette in relazione il fenotipo con il trascrittoma, per cui lo sperimentatore interessato a



CORTESIA D'IRFHNK GENOMICS CONSORTIUM, INDIANA UNIVERSITY

Microarray per lo studio dell'espressione del DNA. Ciascuno dei 16 blocchi esamina in un diverso individuo l'espressione di 96 geni, corrispondenti ai 96 puntini. L'intensità della luminescenza di ciascun puntino indica l'intensità con cui quell'individuo esprimeva il gene corrispondente al momento del prelievo del campione. A destra, particolare di uno dei blocchi.

un comportamento di una pianta (ad esempio la sua resistenza al terreno salino) ne studia il trascrittoma nella speranza di individuare i geni responsabili del fenotipo.

Una diversa, affascinante prospettiva prevede che lo sperimentatore possa simulare l'effetto di un particolare dosaggio di espressione di un set di geni da lui definito e verificare l'effetto di questa simulazione sulla risposta della pianta. Software di questo tipo sono in fase di sviluppo e trovano supporto e impiego nello studio della relazione tra genotipo e fenotipo, ad esempio nei pattern di organizzazione fiorale.

Se le conoscenze a disposizione consentono di predire quale effetto avrà la mancata espressione di un gene sulla tipologia di sviluppo fiorale, sulla base di esperienze sperimentali con l'impiego di mutanti, allora è anche possibile istruire un software al fine di predire situazioni più complesse, magari tramite un algoritmo di autoapprendimento alimentato dai dati di trascrittomica, proteomica e metabolomica, a patto che questi siano corredati da una puntuale e univoca descrizione del materiale vegetale studiato e delle condizioni sperimentali impiegate. La pianta virtuale potrà quindi consentire ai ricercatori di progettare la pianta ideale per ogni possibile condizione ambientale.

L'ottenimento della pianta progettata, tramite le tecnologie genetiche oggi disponibili, consentirà una nuova rivoluzione nell'agricoltura e nelle nume-

rose interfacce tra quest'ultima e i complessi bisogni della comunità umana attuale.

*Pierdomenico Perata, professore ordinario di Fisiologia vegetale alla Scuola Superiore Sant'Anna di Pisa. Amedeo Alpi, professore ordinario di Fisiologia vegetale all'Università di Pisa*

#### Note

1. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408, 796 – 815 (2000).
2. J. Yu *et al.* A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). *Science*, 296, 79 (2002).
3. S. A. Goff *et al.* A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science*, 296, 92 (2002).
4. J.C. Redman, B.J. Haas, G. Tanimoto, C.D. Town. Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. *Plant Journal*, 38, 545-561 (2004).
5. E. Fridman, E. Pichersky. Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Current Opinion in Plant Biology*, 8, 242-248 (2005).
6. O. Thimm, O. Blasing, Y. Gibon, A. Nagel, S. Meyer, P. Kruger, J. Selbig, L.A. Muller, S.Y. Rhee, M. Stitt. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal*, 37, 914-939 (2004).
7. P. Zimmermann, M. Hirsch-Hoffmann, L. Hennig, W. Gruissem. GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox, *Plant Physiology*, 136, 2621-2632 (2004).