

MILLORA GENÈTICA DE PLANTES ASSISTIDA AMB MARCADORS

PERE ARÚS

Centre de Recerca en Agrogenòmica, CSIC-IRTA-UAB.

Adreça per a la correspondència: Pere Arús. IRTA, Centre de Recerca en Agrogenòmica, CSIC-IRTA-UAB. Carretera de Cabrils, km 2. 08348 Cabrils.

Adreça electrònica: pere.arus@irta.es.

RESUM

En els darrers trenta anys s'ha fet un gran progrés en el desenvolupament i ús dels marcadors moleculars en plantes. Els marcadors, que detecten la variació de la seqüència del DNA, permeten la predicció de molts caràcters, com ara la forma, mida, color, resistència a les malalties, productivitat, etc., d'una manera sovint més precoç, eficient i barata que la mesura del caràcter quan aquest es manifesta en la planta. La millora genètica ha anat integrant progressivament els marcadors en el procés de selecció i en el control de qualitat de llavors i planters, de manera que actualment s'han convertit en eines d'ús generalitzat i sovint imprescindible que ajuden i potencien els mètodes clàssics d'obtenció de noves varietats. En aquest article es resumeixen les característiques dels principals tipus de marcadors i es descriuen les seves principals aplicacions en la selecció de gens majors, de caràcters quantitativs i en la incorporació de gens d'interès a partir d'individus exòtics o d'espècies silvestres properes als conreus.

Paraules clau: millora genètica, marcadors moleculars, caràcters quantitativs, introgressió.

MARKER-ASSISTED SELECTION IN PLANT BREEDING

SUMMARY

The last three decades have witnessed an enormous progress in the development and use of molecular markers for plant breeding. Molecular markers are based on the DNA variation and allow the prediction of many characters related with form, color, size, disease resistance, yield, etc., often quicker, more efficiently and cheaper than field measurements when these characters are expressed on the plant. Markers have been progressively integrated in the process of selection and have become a useful and often necessary

tool for breeders to select the best individuals and for quality control of seed and nursery plants. This chapter summarizes the characteristics of the main marker types and their application for the selection of major genes, quantitative characters and for the exploitation of genes of interest from wild or exotic sources in cultivated species.

Key words: plant breeding, molecular markers, quantitative characters, introgression.

INTRODUCCIÓ

La millora genètica de plantes, el procés que porta a l'obtenció de noves varietats d'un conreu, es compon de tres etapes ben diferenciades. En la primera, una vegada definits els caràcters a millorar, es tracta de trobar dins la variabilitat genètica de l'espècie conreada, o de les espècies que s'hi poden hibridar, individus que tinguin aquests caràcters. Aleshores comença la segona etapa, en la qual aquests individus s'hibriden entre si i amb plantes de bones característiques agronòmiques generals, per obtenir una població de base que segregará per a un gran nombre de caràcters, de la qual se seleccionaran els individus que més s'acostin a la varietat que estem cercant, és a dir, els que tinguin un bon comportament general i incorporin els nous caràcters, a la qual sovint s'arriba després de diversos cicles d'hibridació-selecció. El mètode de millora, que consisteix en l'estratègia d'hibridació-selecció i el nombre de cicles necessaris, depèn de molts factors, els més rellevants dels quals són el sistema de reproducció de l'espècie, ja sigui sexual (autògama o al·lògama) o asexual, les característiques genètiques del caràcter (monogènic o poligènic, molt o poc heretable, dominant o recessiu, etc.) i el tipus de varietat que es vulgui obtenir al final (línia pura, híbrid F1, sintètic, clon, multilínia, etc.). Al final d'aquest procés s'arriba a una o més línies seleccionades, de les quals hem de comprovar, dins de la tercera etapa, que són millors en un o més aspectes que les varietats que hi ha al mercat, fet que normalment obliga a realitzar assaigs comparatius. Les línies que millo-

ren les anteriors són les noves varietats que, dins també d'aquesta tercera etapa, s'han de multiplicar, mantenir i comercialitzar per tal que arribin, habitualment com a llavors o plançons, als que les han de conrear.

L'enorme progrés de la genètica molecular des del descobriment de l'estructura del DNA, fa poc més de mig segle, fins ara, ens ha portat a un coneixement cada vegada més detallat del genoma dels éssers vius i al desenvolupament d'un seguit d'eines per al seu estudi, que tenen aplicació en la millora genètica en qualsevol de les tres etapes descrites abans, és a dir, en la caracterització de la variabilitat genètica, en la selecció dels individus amb la composició genètica més interessant en poblacions segregants i, finalment, en el control de la qualitat de les noves varietats, tant en l'aspecte de la seva composició genètica (si cada individu o el conjunt dels que componen una varietat correspon a aquest tipus de varietat), com de la seva identitat (distinció individualitzada de cada varietat). En aquest capítol s'expliquen les característiques i aplicacions d'una d'aquestes eines, els marcadors moleculars i, per extensió, el coneixement del genoma que hem obtingut amb aquests, en la segona de les etapes de la millora, és a dir, en la selecció de noves varietats.

MARCADORS GENÈTICS: DELS GENS MAJORS ALS MARCADORS MOLECULARS

Els marcadors genètics són polimorfismes de qualsevol tipus, que es caracteritzen per tenir una herència mendeliana simple.

Per *polimorfisme* s'entén qualsevol caràcter morfològic o molècula orgànica (incloent-hi, evidentment, les proteïnes i àcids nucleics) que presenti variants al·lèliques que es puguin detectar d'alguna manera. Les propietats ideals dels marcadors són que: *a*) siguin de bona qualitat genètica, és a dir, codominants (que es puguin distingir els dos homozigots i l'heterozigot produïts per qualsevol parella d'allels) i polimòrfics (que cada marcador detecti dos o més allels), *b*) es puguin trobar en gran quantitat en el genoma i el cobreixin completament d'una manera aproximadament uniforme, *c*) no afectin el fenotip de la planta ni interaccionin entre si i *d*) s'obtinguin amb un mètode senzill, robust, barat i adequat pel seu ús a diferent escala. L'objectiu dels marcadors en millora genètica és que ens proporcionin unes fites en el genoma que serveixin per estudiar la seva variabilitat, localitzar gens que estiguin a prop dels marcadors o per poder seguir l'evolució de fragments de DNA, cromosomes o del genoma sencer a través de la seva evolució, en general, o dins d'un procés evolutiu a menor escala, com és la selecció dins d'un programa de millora genètica.

Els primers marcadors genètics que es van usar foren els que determinen les característiques qualitatives que afecten qualsevol aspecte de la morfologia o fisiologia de la planta, com la forma, mida o color del vegetal o qualsevol dels seus òrgans, la seva resistència a malalties, les èpoques de floració o fructificació, etc. Aquests marcadors, alguns dels quals codifiquen caràcters de gran importància econòmica, i que s'anomenen també *gens majors*, són tan antics com el mateix naixement de la genètica, ja que van ser la base de la formulació de les lleis de l'herència de Mendel a mitjan segle XIX. Tenen com a limitacions el seu baix nombre en la majoria d'espècies, la seva forta interacció amb altres gens (que fa difícil que se'n puguin estudiar molts en segrega-

ció en el mateix encreuament), el seu caràcter habitualment dominant, el poc nivell de polimorfisme (rarament més de dos allels) i la condició freqüentment deletèria d'alguns allels, normalment els recessius. L'estudi de molts d'aquests gens al mateix temps en descendències d'encreuaments controlats (de tipus F_2 , la descendència obtinguda de l'autofecundació de l'híbrid entre dues línies diferents, o retroencreuament (RC), la descendència que s'obté d'encreuar aquest híbrid amb un dels seus pares) va fer possible la construcció dels primers mapes genètics, que representen l'alineació ordenada d'aquests marcadors en els cromosomes basada en el grau de lligament entre marcadors. La construcció de mapes genètics amb aquests marcadors fou una empresa a llarg termini, plena de dificultats i reservada només a les espècies més estudiades des del punt de vista genètic com el blat, el blat de moro, el tomàquet o el pèsol (O'Brien, 1990), en les quals aquesta inversió podia justificar-se per la seva qualitat d'espècies model per als estudis genètics.

Pels tres quarts del segle passat es va descobrir una nova classe de marcadors, els isoenzims, que havia de revolucionar la genètica clàssica i assentar les bases dels usos actuals dels marcadors (Lewontin, 1970; Tanksley i Orton, 1983). Els isoenzims són formes moleculars d'un mateix enzim amb diferències moleculars d'origen genètic (de la seqüència del gen que els codifica) que en permeten la separació per electroforesi (la substitució d'un sol aminoàcid pot produir una migració electroforètica suficientment diferent per ser detectada). Els isoenzims permeten identificar diferents formes al·lèliques d'un enzim determinat que sovint s'expressen de manera codominant. Els isoenzims polimòrfics solen detectar sèries al·lèliques i és possible comparar la variabilitat dels mateixos isoenzims en diferents espècies, cosa que els fa especialment útils per als estudis de variabilitat genètica

comparativa (Gottlieb, 1981). La metodologia és senzilla i barata, però el seu baix nombre (és infreqüent que s'hagin descrit més d'una o dues desenes d'isoenzims polimòrfics en les espècies en les quals s'han usat) és el seu factor limitant fonamental. Els isoenzims van permetre l'anàlisi de la variabilitat genètica des d'una perspectiva més acurada i profunda (Lewontin, 1970; Gottlieb 1981), van ser una tècnica que va expandir aquests estudis de variabilitat a moltes espècies silvestres i cultivades i van fer possible els primers assaigs de les aplicacions que després han estat profusament usades quan altres marcadors més potents han estat disponibles (Tanksley i Orton, 1983).

Quan l'ús dels isoenzims s'havia expandit considerablement, va aparèixer un nou tipus de marcador, els RFLP (*restriction fragment length polymorphisms*) (Botstein *et al.*, 1980), que foren els primers directament basats en la variació del DNA i els primers que s'anomenaren *moleculars*. Els RFLP es basen en les diferències de la mida de fragments homòlegs de DNA creats a partir de la digestió del DNA total amb un enzim de restricció. Aquests polimorfismes són deguts a la detecció o no-detecció d'una diana de l'enzim en un fragment determinat, o a l'existència de diferències de mida entre fragments al·lèlics delimitats per les mateixes dianes de restricció, a causa de l'existència d'una inserció (o deleció) dins d'un dels dos al·lels. Els RFLP es detecten amb el mètode de transferència Southern (Southern, 1975), consistent a separar els productes de la digestió per electroforesi, fixar-los en una membrana de niló i identificar-los allí amb sondes curtes de DNA marcades. Amb els RFLP es pot produir un nombre virtualment infinit de marcadors, i per això superen la principal limitació dels isoenzims i obren la porta als mapes *saturats*, és a dir, amb una cobertura completa del genoma amb marcadors. El primer ma-

pa en plantes va ser el del tomàquet (Bernadsky i Tanksley, 1986) al qual van seguir molts altres en espècies econòmicament importants. Els RFLP són de gran qualitat genètica (codominants, polimòrfics) i altament transferibles entre espècies (normalment una sonda d'una espècie és útil per produir RFLP en qualsevol membre de la seva família i freqüentment en la d'espècies d'altres famílies). Això els ha fet els més utilitzats per a la comparació de mapes, en què la posició de marcadors homòlegs de diferents espècies permet establir el nivell de semblança entre els seus genomes, de manera que s'identifica la seva *sintènia*, o localització dels mateixos marcadors als mateixos cromosomes, i la seva *colinearitat*, o manteniment de l'ordre entre marcadors sintènics. La comparació de genomes és un element important en la comprensió de la seva evolució (Ahn i Tanksley, 1993; Devos, 2005; Yogeewaran *et al.*, 2005) i en la predicció de l'existència de gens d'interès en espècies poc estudiades comparant el seu genoma amb el d'altres en les quals s'ha acumulat un major nivell de coneixement (Ku *et al.*, 2001; Barnes, 2002). Tot i les seves indubtables virtuts, el principal defecte dels RFLP és que s'obtenen amb un mètode feixuc, lent i car, només a l'abast d'alguns laboratoris especialitzats, adequat per a estudis fonamentals, però lluny d'usos a gran escala com els necessaris per a la seva aplicació en millora genètica.

El descobriment de l'amplificació (producció d'un elevat nombre de còpies) de fragments curts de DNA basada en el mètode de la reacció en cadena de la polimerasa (PCR) (Saiki *et al.*, 1988), impulsà el desenvolupament d'una nova onada de marcadors. La PCR permet l'amplificació exponencial de qualsevol fragment de DNA petit (100-2.000 pb) a partir d'una sèrie de cicles de síntesi (habitualment entre trenta i cinquanta) de noves cadenes de DNA, que es fa amb l'ajut d'un enzim termostable (la po-

limerasa *Taq*) i dos encebadors (fragments de DNA, freqüentment de 18-24 bases, complementaris de la seqüència de cada extrem del fragment a amplificar). Aquest fragment amplificat es pot separar per electroforesi, i llavors es pot comparar la seva mida amb la d'altres fragments o es pot aïllar per a la seqüenciació posterior. Els primers marcadors basats en la PCR foren els RAPD (*randomly amplified polymorphic DNA*) (Williams *et al.*, 1990), en què la PCR es fa amb un sol encebador molt curt (~10 bases) que amplifica per atzar els fragments que aquest encebador pot delimitar, i produeix normalment un nombre discret (~2-20) de fragments amplificats, que es poden separar per electroforesi. La presència o absència dels fragments amplificats en una descendència o col·lecció d'individus permet l'anàlisi genètica d'aquests polimorfismes i l'estudi de la seva variabilitat. Poc després dels RAPD es van proposar els AFLP (*amplified fragment length polymorphisms*) (Vos *et al.*, 1995), en què la variabilitat es crea, tal com en els RFLP, amb la digestió del DNA total, en aquest cas amb dos enzims de restricció. Posteriorment s'afegeixen als extrems dels fragments obtinguts uns adaptadors (fragments curts de 14-20 pb de seqüència coneguda) i es redueix el nombre de fragments a força d'amplificar el DNA usant uns encebadors homòlegs a la seqüència dels adaptadors que incorporen un nombre petit de bases noves a l'atzar (entre una i tres) que, en proporció inversa a aquest nombre, amplifiquen més o menys fragments. Finalment els fragments són separats per electroforesi i la seva variabilitat és analitzada en un conjunt d'individus.

RAPD i AFLP, així com altres marcadors que s'han desenvolupat al voltant de variacions sobre els mateixos mètodes (Garcia-Mas *et al.*, 2000), tenen en comú la seva capacitat de produir una gran quantitat de fragments amplificats potencialment polimòrfics i, per tant, de marcadors que s'obtenen d'una ma-

nera *inespecífica*, és a dir, sense tenir cap informació prèvia sobre la seva seqüència. La propietat de generar molts polimorfismes a partir d'un mètode senzill va ser crucial en l'acceptació d'aquests marcadors, que s'usen àmpliament en estudis de variabilitat, en la construcció de mapes i en la cerca de marcadors prop de gens d'interès, com comentarem més endavant. El seu ús va convertir el mapatge en una operació més senzilla, ràpida i barata del que havia estat amb els RFLP, i contribuï a l'extensió de l'ús dels marcadors a moltes més espècies. Malgrat tot, tant RAPD com AFLP tenen punts febles molt importants, ja que són marcadors generalment dominants i poc transferibles i, en el cas dels RAPD, són poc reproduïbles. Com a conseqüència, la informació que han proporcionat, tant en mapes com en variabilitat, es limita a la que s'ha obtingut en les descendències o col·leccions de genotips estudiats, sense que aquesta informació sigui gairebé mai acumulable o ampliable amb la que proporcionen altres marcadors, altres genotips o altres poblacions de mapatge. El seu ús per a aquests propòsits ha decaït considerablement amb l'aparició de marcadors de millor qualitat, i és previsible que aquesta davallada continuï en el futur.

Els problemes de transferibilitat, reproductibilitat i, de vegades, de qualitat (dominància) de RAPD i AFLP es poden resoldre si és possible aïllar el fragment de DNA polimòrfic i convertir-lo en un marcador de PCR *específic*, és a dir, basat en la seqüència d'aquest fragment, a partir de la qual es dissenyen un parell d'encebadors que permeten amplificar-lo. Aquesta és una operació d'una certa complexitat que sovint, encara que no sempre, es pot realitzar amb èxit, però que es limita a aquells marcadors que tenen un valor especial, normalment perquè estan lligats a gens d'interès. Aquests principis van portar al desenvolupament de marcadors de PCR elaborats a partir de la seqüència de qualsevol frag-

ment de DNA (com per exemple la sonda d'un RFLP), que s'anomenen d'una manera general STS, (*sequence tagged sites*) (Olsen *et al.*, 1989), o SCAR (*sequence characterized amplified regions*), en el cas que s'obtinguin de RAPD (Paran i Michelmore, 1993). Una elaboració dels SCAR o STS són els marcadors CAPS (*cleaved amplified polymorphic sequence*) (Akopyanz *et al.*, 1992), en què a partir del polimorfisme de seqüència de dos allels d'un RAPD, AFLP o qualsevol altre fragment de DNA, es pot elaborar un marcador de PCR i buscar un enzim de restricció que talli un dels allels però no l'altre, cosa que genera un polimorfisme, específic i habitualment codominant.

El progrés en el coneixement de la seqüència del genoma dels éssers vius, degut en part al desenvolupament, al final del segle xx, de tecnologies eficients de seqüenciació de DNA, resultaren en el descobriment d'un dels marcadors més potents i versàtils dels que s'usen actualment. Es tracta dels marcadors basats en microsatèl·lits, anomenats també SSR (*simple-sequence repeats*). Un microsatèl·lit és una seqüència curta de DNA formada per un motiu d'una a sis bases repetit diverses vegades en tàndem. Els microsatèl·lits, que són molt freqüents i es troben dispersos per tot el genoma, tant en zones transcrites com no transcrites del DNA, tenen la particularitat de ser altament polimòrfics a causa de la deleció o inserció d'un o més dels motius dels quals es compon (Morgante i Olivieri, 1993). Aquestes propietats els fan adequats per convertir-los en marcadors de PCR quan es coneixen seqüències que els continguin i permetin dissenyar encebadors en les zones de seqüència única que flanquegen cada microsatèl·lit. El marcador (SSR) resultant detectarà habitualment un sol locus codominant i altament polimòrfic. El mètode d'anàlisi és senzill, escalable, requereix molt poc DNA, no necessàriament de bona qualitat, habitualment robust, i té el cost d'una simple

reacció de PCR i la separació posterior per electroforesi, és a dir, comparativament barat. Els SSR solen ser altament transferibles dins de l'espècie, mentre que un percentatge bastant elevat ho és entre espècies properes del mateix gènere, i molt més baix entre gèneres propers. Això permet acumular informació respecte a la variabilitat genètica entre dades de diferents laboratoris o col·leccions de genotips, o inferir la posició d'un marcador situat en un mapa respecte d'altres mapes de la mateixa espècie. En conjunt, les propietats dels SSR s'acosten molt a les del marcador ideal. No obstant això, la primera dificultat per al seu ús va ser la disponibilitat de seqüències de DNA per poder cercar-los, que només eren suficients en les espècies més estudiades. Aquesta limitació no és actualment un problema en la majoria d'espècies conreades, que, amb el desenvolupament, primer de genoteques enriquides en microsatèl·lits (Billotte *et al.*, 1999), i després de les col·leccions de fragments de DNA transcrits (EST, *expressed sequence tags*), va quedar resolt en gran part (Morgante *et al.*, 2002). Actualment s'han desenvolupat, o hi ha la informació per fer-ho, centenars o milers de SSR en molts conreus. Els SSR són actualment els marcadors més usats en l'àmbit aplicat i, per bé que tard o d'hora s'aniran substituint per altres marcadors, tot sembla indicar que seguiran sent útils durant bastant temps.

En els darrers anys s'han començat a usar marcadors basats en un dels darrers nivells de la variabilitat genètica, la produïda per la substitució d'una base per una altra. Aquests marcadors, que s'anomenen SNP (*single nucleotide polymorphisms*), són també la conseqüència de la informació que es va acumulant sobre la seqüència del DNA en moltes espècies i del fet que el nombre de genomes de plantes complets o parcialment seqüenciats és cada vegada més gran. Molts dels marcadors que hem considerat abans, tots menys els SSR, ja tenien com una pos-

sible base de variabilitat els SNP, que es poden detectar amb tecnologies molt diverses, com la reseqüenciació de fragments de DNA de polimorfisme conegut, els mètodes basats en la piroseqüenciació (Alderborn *et al.*, 2000), l'espectrometria de masses (Rodi *et al.*, 2002), l'HPLC desnaturalitzant (O'Donovan *et al.*, 1998), la PCR Taq-man (Livak, 1999) i l'ús de microxips (Borevitz *et al.*, 2003), entre d'altres. Cadascuna d'aquestes tecnologies s'adapta a diferents situacions considerant dos paràmetres bàsics: el nombre d'individus a analitzar amb el mateix SNP i el nombre de marcadors a analitzar per cada individu. D'una manera general, els SNP s'analitzen com a bial·lèlics, codominants o dominants i el seus avantatges són que, en ser els marcadors més ubics, és possible trobar-ne en qualsevol punt del genoma i, si s'analitzen a gran escala, el cost per dada individual pot ser molt inferior al de qualsevol altre marcador. La seva aplicació depèn del nivell del coneixement de la seqüència de DNA de l'espècie d'interès, fet que els limita a les poques espècies en les quals aquesta informació és molt gran. És previsible, doncs, que el seu ús s'estengui a mesura que la informació sobre la seqüència de DNA augmenti, tal com ha passat amb els SSR.

Recentment s'han posat en el mercat uns aparells de seqüenciació massiva de DNA (www.rock.com, www.illumina.com, www.appliedbiosystems.com), que permeten seqüenciar un gran nombre de fragments de DNA més o menys grans (de 25 a 400 pb), amb una quantitat total entre un quart i una megabase per assaig, i amb perspectives que aquests nombres augmentin ràpidament. Aquests aparells, que ja estan facilitant la seqüenciació de genomes *de novo*, seran molt útils per a la seva reseqüenciació (obtenció de la seqüència d'un altre individu d'una espècie una vegada que ja n'hi ha una de completa, que és més senzill que l'obtenció de la primera seqüència, que re-

quereix un procediment d'obtenció i muntatge més lent i costós) i es poden convertir en eines eficients per a l'anàlisi de l'expressió gènica. També poden representar una millora substancial en el descobriment de SNP nous o altres tipus de polimorfismes (derivats de la inserció o deleció de petits fragments de DNA). L'anàlisi de genotips heterozigots o de barreges de DNA total o de fragments transcrits de diversos individus pot generar un gran nombre de SNP en poc temps i a preus competitius (Barbazuk *et al.*, 2007), fet que expandirà l'ús d'aquests marcadors a moltes més espècies i previsiblement marcarà el futur immediat de les aplicacions dels marcadors en la millora genètica.

SELECCIÓ ASSISTIDA AMB MARCADORS

Dins d'aquesta secció s'inclouen diverses aplicacions dels marcadors com a ajut al procés de selecció de varietats noves. Aquestes aplicacions estan dirigides a l'augment de l'eficiència de la selecció de caràcters determinats en el sentit de fer-la més precoç, més acurada o més barata. Els marcadors permeten també la selecció d'individus amb un genotip interessant quan el fenotip seria difícil o impossible d'avaluar (heterozigots per a gens recessius, acumulació de gens de resistència a una mateixa malaltia), i faciliten la selecció del genoma d'un genotip o una espècie en programes d'introducció (*introgressió*) de gens d'espècies silvestres o genotips exòtics. Amb els marcadors també és possible conèixer a fons l'herència dels caràcters agronòmics, especialment aquells que són més difícils, com els d'herència quantitativa, i fa més eficaç, a partir d'aquesta comprensió, la selecció dels genitors que tinguin els al·lels adequats i dels descendents amb les combinacions de gens més interessants.

Etiquetatge de gens

La meiosi produeix sobreencreuaments d'una manera discreta en el genoma de les plantes. El nombre de recombinacions de cada cromosoma després de la meiosi és petit. És freqüent que no n'hi hagi cap, o bé un o dos, i nombres majors són progressivament més rars, encara que això depèn de la mida del cromosoma, de l'espècie i de l'individu. Per tant, els fragments que es conserven de l'individu original en el seu descendent són habitualment grossos, cosa que vol dir que un allel d'un marcador que estigui a la vora d'un gen d'interès s'heretarà habitualment amb un allel d'aquest mateix gen. La freqüència amb la qual això passi dependrà de la distància genètica entre el marcador i el gen, un paràmetre que es calcula a partir de la freqüència de recombinació entre el gen i el marcador, que és la informació bàsica necessària per a la construcció de mapes de lligament. La distància genètica es pot determinar en les descendències típicament usades per al mapatge (de tipus F_2 o retroencreuaments), sabent el fenotip del gen i del marcador per a cada individu. Es mesura en centimòrgans (cM) —per donar una idea del seu valor, per a distàncies curtes (1-10 cM), un cM és aproximadament igual al percentatge de recombinació (1-10 %). Quan un gen que ens interessa és a prop d'un marcador, podrem seleccionar el marcador en comptes del gen, com si tinguéssim el nostre gen etiquetat amb el marcador i el poguéssim reconèixer amb aquesta etiqueta. Com que els dos locus segreguen en la descendència, això vol dir que hi haurà un allel del marcador que estarà associat a l'allel del gen que ens interessa i un altre que ho estarà de l'allel del gen que volem eliminar. L'eficiència de la selecció amb el marcador dependrà de la proximitat entre els dos locus: si la distància és de 5 cM, ens equivocarem aproximadament el 5 % de les vegades, si és d'1 cM, l'u per cent, i si el gen

i el marcador cosegreguen ~0 cM (fet que passa quan el marcador està basat en la seqüència del mateix gen), no ens equivocarem gairebé mai.

Un dels avantatges de l'etiquetatge de gens és que es pot fer la selecció tan aviat com sigui possible extreure el DNA de la planta, cosa que en general ocorre molt aviat (la llavor, el cotiledó, la primera fulla vertadera). Si el gen que ens interessa es manifesta molt més tard, com ara un caràcter de flor o fruit, requereix una tecnologia addicional per manifestar-se, com en el cas de la major part de resistències a malalties o plagues, o hi és però no pot manifestar-se (un allel recessiu en segons quin tipus d'encreuament); llavors la selecció amb el marcador pot ser avantatjosa i pot representar un estalvi de temps, diners o eficiència respecte a la selecció basada en el fenotipus. També, la selecció basada en el marcador només descarta les plantes que tenen un cert caràcter, però permet seleccionar d'una manera convencional la resta. La selecció es pot aplicar al mateix temps en tots els gens en què la població estigui segregant per als quals es disposi de marcadors, i es pot fer per a diversos gens que afectin el mateix caràcter, encara que una vegada que un s'hagi incorporat no es pugui notar cap canvi en el fenotip. Aquest és el cas de molts gens de resistència a malalties en què la seva acumulació (que s'anomena també *piramidació*) no canvia la resistència original de la planta, però pot representar un avantatge pel que fa a la seva durada (Castro *et al.*, 2003; Tar'an *et al.*, 2003; Barloy *et al.*, 2007). La piramidació només es pot fer amb l'ajut dels marcadors, ja que sense aquests no es pot tenir cap garantia que el segon gen o els posteriors s'hagin incorporat al genoma de la planta seleccionada.

Com es poden trobar marcadors a prop de gens d'interès?

Per poder etiquetar gens cal trobar marcadors propers, idealment molt a prop, perquè d'això depèn l'eficiència de la selecció. Hi ha diverses maneres de fer-ho. Una consisteix a construir un mapa genètic amb una població que segregui per al caràcter d'interès, de manera que es pugui trobar així la posició del gen. Quan en l'espècie existeix ja un mapa saturat amb marcadors transferibles, una bona manera de procedir és seleccionar alguns marcadors d'aquest mapa, de manera que cobreixin la major part del genoma a distàncies més o menys grans (25-35 cM) i que segreguin la població usada per mapar el gen. Una vegada la posició del gen queda establerta amb aquest mapa mínim, s'assagen altres marcadors situats en la regió genòmica on hi ha el gen fins que se'n troba un que està suficientment a prop (< 2 cM) del gen. Si no hi ha marcadors suficientment propers es poden intentar cercar dos marcadors que flanquegin el gen, encara que estiguin a distàncies majors: seleccionant al mateix temps dos marcadors d'aquestes característiques separats entre si, per exemple, 10 cM, s'aconsegueix una eficiència igual o major que amb un marcador que estigues a 0,25 cM (Tanksley, 1983).

El mètode més usat per a la localització de marcadors propers a un gen, anomenat *bulked segregant analysis* (BSA), no requereix la construcció de cap mapa (Michelmore *et al.*, 1991). Es basa en la característica ja esmentada que en les descendències habitualment usades en els experiments de mapatge, el nombre de sobrecreuaments per cromosoma i meiosi és baix, i per tant, l'al·lel que cerquem del gen d'interès s'hereta juntament amb una gran fragment de DNA d'un dels genitors, mentre que l'al·lel alternatiu d'aquest gen inclou un fragment també gran, de DNA homòleg, però diferent, procedent de l'altre genitor. D'aquesta ma-

nera, si en un encreuament determinat (suposem una F_2) barregem el DNA de plantes (normalment entre sis i dotze) amb un dels dos fenotips (per exemple, resistent a una malaltia) i el comparem amb una barreja semblant de DNA de plantes amb el fenotip alternatiu (susceptible) per a un gran nombre de marcadors, si trobem algun marcador que detecti un polimorfisme entre les dues barreges, molt probablement es trobarà en la regió on hi ha el gen que busquem. Això és degut al fet que en cada barreja només al voltant del gen seleccionat es mantindrà una finestra en la qual el DNA serà només d'un dels dos possibles haplotips (fragments al·lèlics). La mida d'aquesta finestra serà més petita com més gran sigui el nombre de plantes usat en cada barreja. Els marcadors més adequats per a aquests assaigs són els RAPD, AFLP i els marcadors inespecífics semblants, atesa la seva elevada capacitat de generar polimorfismes de DNA. El BSA és un mètode de gran eficiència que ha estat sovint utilitzat amb èxit. Com a punts febles, hi ha la seva gran sensibilitat als errors de fenotipatge que puguin implicar la inclusió d'alguna planta de fenotip erroni en alguna de les barreges de DNA, així com de plantes amb fenotip correcte no pertanyents a l'encreuament, que s'han de comprovar amb marcadors abans de ser usades.

Un altra via possible de localització de marcadors es basa en la comparació de genomes, quan l'espècie d'interès és propera a una altra de la qual hi ha molta informació sobre la seqüència, o de la qual fins i tot es coneix la seqüència sencera del genoma. En aquests casos, una vegada localitzada la regió on es troba al gen a seleccionar, es poden cercar en la regió homòloga de l'espècie model fragments de DNA que puguin generar marcadors lligats al gen d'interès (Ku *et al.*, 2001).

El millor marcador és el gen mateix

A mesura que es va avançant en el coneixement dels gens, les seves característiques i funcions i la seva posició en els genomes, és possible dissenyar estratègies que permetin localitzar els gens d'interès i elaborar marcadors a partir de la seva seqüència, que reemplaçarien els marcadors propers al gen. Això té molts avantatges, i el primer és que la selecció amb un marcador basat en la seqüència del mateix gen és la més acurada de totes les possibles.

Una estratègia per cercar un gen d'interès és el clonatge posicional, que es basa a trobar, per la via del BSA o d'altres semblants, marcadors que estiguin en la proximitat del gen d'interès i procedir després al mapatge d'alta resolució d'aquests marcadors amb un nombre molt elevat (generalment milers) de plantes (Tanksley *et al.*, 1995). Una vegada hi ha dos marcadors que flanquegen el gen a una distància molt curta (< 1 cM), es poden usar per garbellar genoteques (col·leccions de grans quantitats de fragments de DNA) de mida gran (habitualment BAC, *bacterial artificial chromosomes*, que són fragments de ~100-250 kb) de l'espècie en què s'estudia el gen. Si els dos marcadors es troben en un mateix BAC, s'ha de seqüenciar aquest; si no, s'ha de construir un *contig* (conjunt de BAC parcialment sobreposats) usant marcadors obtinguts de la seqüència dels BAC (habitualment dels extrems) que contenen els marcadors, fins a localitzar un BAC que contingui el gen, és a dir, un BAC amb el qual es puguin obtenir dos marcadors, un a cada cantó d'aquest. Una vegada seqüenciat el BAC, s'identifiquen, entre els gens que conté, aquells que corresponen més probablement al gen cercat. Experiments de complementació per la via transgènica o altres aproximacions permetran al final validar quin d'aquests gens és el causant del fenotip estudiat. Aquest mètode ha permès clonar gens molt importants sobre els

quals no es tenia cap informació sobre la seva base molecular, com per exemple els primers que es van conèixer dels que confereixen resistència a malalties (Martin *et al.*, 1993; Cai *et al.*, 1997). Es tracta, però, d'una estratègia complexa, que implica freqüentment una inversió considerable de temps i mitjans i no està exempta de riscos: per exemple, que la regió estudiada estigui en una zona de baixa recombinació, on petites distàncies genètiques puguin correspondre a grans distàncies físiques, o que els dos al·lels alternatius es trobin en una regió poc variable, que faci difícil la localització de marcadors.

La disponibilitat de mapes físics, és a dir, basats en la distància física (nombre de bases) i no en la distància genètica (freqüència de recombinació), com els mapes de lligament, simplifica molt el clonatge de gens (Zhang i Wing, 1997; Chang *et al.*, 2001). Els mapes físics es fan sovint amb l'acoblament de grans fragments de DNA procedents de genoteques de BAC en *contigs* que cobreixin la major part o la totalitat del genoma. Quan es disposa d'un mapa físic, el clonatge de gens localitzats amb una precisió raonable al mapa genètic requereix essencialment la cerca de BAC que continguin marcadors a prop d'aquest gen i l'aproximació successiva al gen amb marcadors obtinguts de BAC adjacents a aquest fins trobar-ne un que contingui el gen que busquem. La seqüenciació d'aquest BAC ens durà a la identificació de candidats i finalment al gen. Quan el genoma està complet o parcialment seqüenciat, i una vegada situat el gen en una regió del genoma, serà suficient la cerca *in silico* de gens d'aquesta regió que puguin correspondre per la seva funció al gen que cerquem, la comprovació de la seva cosegregació amb el gen i la validació de la seva funció via complementació o altres aproximacions.

Una aproximació funcional a la cerca de gens d'interès és l'anàlisi de gens candidats

(GC). Un GC és un gen de seqüència coneguda, la funció hipotètica o demostrada del qual pot estar implicada en l'expressió d'un caràcter determinat. El mapatge de GC en el genoma d'una planta i la seva colocalització amb la de gens que afectin aquest caràcter ens pot permetre establir una relació de causa-efecte entre els GC i aquests gens. L'estratègia dels GC ha estat usada amb èxit en molts casos (Pflieger *et al.*, 2001), però bé que en altres no ha donat resultat. Les causes dels fracassos són que per a alguns caràcters és difícil cartografiar tots els GC potencialment involucrats en el caràcter (no hi ha seqüències disponibles en l'espècie, no és possible trobar polimorfisme en tots els GC assajats, etc.) i que de vegades pot ser difícil identificar tots els possibles GC perquè corresponen a àmbits funcionals poc coneguts o a mecanismes de regulació gènica (factors de transcripció, mRNA, etc.). Una via envers el reconeixement del màxim possible de GC és l'ús de microxips, constituïts per milers de fragments de DNA transcrits immobilitzats en un xip, placa de vidre o altres materials, que s'hibriden amb DNA marcat procedent de la part transcrita d'un teixit tractat i un d'usat com a control per estudiar l'expressió diferencial de gens entre si. Els gens que s'expressen diferencialment en tractaments, teixits o moments del desenvolupament relacionats amb l'expressió del caràcter estudiat són també candidats a explicar la variació fenotípica estudiada, i representen generalment una mostra molt àmplia del genoma transcrit, difícil de detectar per altres mitjans, que augmenta l'amplitud i probabilitats d'èxit d'aquest tipus d'anàlisi.

Caràcters d'herència quantitativa

La major part dels trets agronòmics (relacionats amb la productivitat, la qualitat del producte, l'adaptació de la planta a di-

ferents medis, etc.) són de tipus quantitatiu, és a dir, estan codificats per diversos gens, l'expressió dels quals està influïda també per l'ambient i per la interacció entre els gens i l'ambient. Durant molts anys es van analitzar acceptant el que s'anomenà *hipòtesi infinitesimal*, que assumeix que la base genètica d'aquests caràcters està determinada per molts gens amb efectes individuals petits i semblants. El descobriment dels marcadors moleculars i la generació de mapes genètics amb aquests va permetre re-examinar aquests caràcters, incloent-hi una dimensió addicional: la posició dels gens al genoma. L'anàlisi de la cosegregació entre marcadors que cobreixen tot el genoma i les dades dels caràcters quantitatius permeten establir punts del genoma on existeix una correlació entre les dades d'alguns marcadors i les dades fenotípiques. Es pot inferir que en aquests llocs es troben un o més gens que afecten l'expressió del caràcter quantitatiu, que s'anomenen QTL (*quantitative trait loci*). La diferent intensitat d'aquesta correlació i els efectes produïts per homozigots i heterozigots per als marcadors en la proximitat del QTL permeten també establir algunes de les característiques centrals d'aquests QTL, és a dir, la magnitud del seu efecte, el seu mode d'expressió (additivitat, dominància, sobredominància, etc.) i la seva posició aproximada en el mapa genètic. També és possible estimar la interacció entre diferents QTL (Ahmadi *et al.*, 2001) i entre aquests i el medi ambient (MacMillan *et al.*, 2006).

La prova de concepte que aquesta aproximació era viable i generava informació útil va ser proposada per primera vegada en plantes per Tanksley *et al.* (1982), quan els marcadors que hi havia eren els isoenzims. Posteriorment, Paterson *et al.* (1988), ja amb un mapa del tomàquet saturat, i construït en la seva major part amb RFLP, van realitzar el primer estudi de QTL amb cobertura de tot el genoma i amb una anàlisi estadís-

tica específicament adequada als estudis de lligament. Els resultats desmentiren clarament la hipòtesi infinitesimal i demostraren (Tanksley, 1993) que una proporció, variable en funció de molts aspectes (espècie, caràcter, població analitzada i les seves característiques, etc.) de la variància fenotípica de cada caràcter està determinada per un nombre variable i finit de QTL, amb magnituds de l'efecte i accions gèniques molt diferents entre si. Aquests resultats obriren la porta a l'ús de marcadors propers a aquests QTL per a la selecció d'aquests caràcters, de la mateixa manera que es fa amb els caràcters qualitius. En alguns casos l'anàlisi de QTL porta al descobriment d'un o dos QTL de gran efecte que poden tractar-se d'una manera independent, tal com si fossin gens majors, i dels quals es poden buscar i trobar diferents al·lels. També és possible clonar QTL per la via dels gens candidats o amb aproximacions posicionals (Friedman *et al.*, 2000).

Els desenvolupaments genòmics en els àmbits de la transcriptòmica (anàlisi del DNA transcrit i de la seva expressió diferencial en teixits, ambients o genotips), la proteòmica (anàlisi de la composició, estructura i expressió de les proteïnes) i la metabolòmica (estudi bioquímic de metabòlits i polipèptids i les vies metabòliques que els produeixen), que impliquen en tots els casos l'anàlisi massiva i sovint quantitativa dels gens i els seus productes, obren una via nova a l'estudi del fenotip, molt més propera al missatge genètic que l'avaluació de caràcters morfològics i fisiològics, i complementària a aquesta. Els resultats de les anàlisis instrumentals en transcriptòmica, que sovint es fan amb l'ajut de microxips, consisteixen en dades quantitatives de presència relativa de milers de seqüències. Aquestes dades (i també les produïdes per assaigs semblants de proteòmica i metabolòmica) es poden tractar com qualsevol altre caràcter fenotípic en una població

segregant i detectar la seva correlació amb la segregació de marcadors que cobreixen tot el genoma. Aquesta aproximació, que s'anomena *genòmica genètica* (Jansen i Nap, 2001), permet la detecció de QTL (que en aquest cas s'anomenen eQTL o QTL d'expressió) que identifiquen regions del genoma on hi ha gens involucrats en resposta a perfils d'expressió determinats. Les posicions dels eQTL poden ser comparades amb les de gens majors o QTL que determinen diferències morfològiques o agronòmiques, fet que en alguns casos permet la identificació o validació de gens candidats.

Tot i que aquestes noves aproximacions són molt útils i d'aplicació immediata en alguns casos, els caràcters quantitius segueixen presentant dificultats per a l'anàlisi en molts d'altres, particularment en caràcters habitualment poc heretables, per als quals la majoria dels QTL detectats són de petit efecte, i la seva expressió és fortament afectada pel medi ambient, com els relacionats amb la productivitat en els cereals i altres espècies de conreu extensiu. Un dels elements que dificulta l'anàlisi genètica d'aquests caràcters és que les poblacions usades, habitualment del tipus F_2 , RC, RIL (línies recombinants consanguínies) o DHL (línies diplohaploides), són poc eficients per a aquest propòsit, ja que cada individu és una barreja d'al·lels de QTL de cadascun dels seus pares que interaccionen entre si i dificulten la individualització i avaluació de l'efecte de cada QTL i de la seva acció gènica. El nombre d'individus necessari per a una bona avaluació del caràcter hauria de ser elevat (diversos centenars) i cada individu s'hauria d'estudiar en diferents replicacions i sotmetre's a medis diferents. Algunes d'aquestes poblacions són poc adequades per fer-ne replicacions, com les F_2 o RC. Una via a la resolució d'aquests problemes és l'anàlisi de generacions avançades de retroencreuaments (Tanksley i Nelson, 1996) i la creació de col·leccions de línies

quasiisogèniques (NIL) (Eshed i Zamir, 1995), cadascuna de les quals conté el genoma sencer d'un dels genitors de l'encreuament inicial (idealment una línia pura, la qual anomenarem *genitor recurrent* o R) a excepció d'un fragment d'un dels cromosomes, que pertany a l'altre (el *genitor donant* o D). La selecció de les NIL es fa usant marcadors situats per tot el genoma en diferents generacions de retroencreuament i autofecundació a partir d'un híbrid $R \times D$. Si es disposa d'una col·lecció de NIL diferents que cobreixin la totalitat del genoma de D, qualsevol diferència entre un caràcter d'una de les NIL i R pot ser atribuïble a l'existència d'un gen o QTL localitzat en el fragment introgredit. Encreuant R amb qualsevol NIL es pot obtenir un heterozigot que ens permet avaluar l'efecte de diferents dosis del mateix allel sobre el mateix fons genètic, o encreuant diferents NIL podem estimar l'efecte conjunt de diversos QTL del mateix caràcter i les seves interaccions.

Mapatge per associació

El mapatge per associació o per *desequilibri de lligament* (DL) és una altra manera de localitzar gens d'interès o marcadors per a aquests (Flint-Garcia *et al.*, 2003; Gupta *et al.*, 2005). El DL és l'associació no a l'atzar d'allels de dos locus, de manera que un allel determinat del primer locus prediu la presència d'un altre allel del segon. Tal com s'ha explicat en la descripció del BSA, els fragments de DNA conservats en una població de mapatge són grans, cosa que vol dir que trobarem associació (o DL) entre marcadors que estiguin localitzats a distàncies grans del mateix cromosoma. Això vol dir que la nostra capacitat d'estimar distàncies petites serà limitada o ens obligarà a usar un nombre de plantes molt gran. Les col·leccions de genotips poc relacionats, en què el nombre de recombinacions que s'han produït en el

seu genoma des d'un ancestre comú és molt elevat, tenen un nivell de DL molt més baix, que pot reduir els fragments conservats fins a mides molt petites (centenars o milers de bases). Aquestes característiques fan que si trobem associació en una col·lecció de genotips amb DL baix entre un marcador i un gen, aquest marcador estarà molt a prop del gen, o serà part de la seqüència d'aquest. Això és útil per validar gens candidats, col·localitzats amb gens o QTL d'una manera aproximada amb el mapatge convencional, i per als quals el manteniment de l'associació amb el caràcter en col·leccions de genotips proporciona una prova important de la seva relació amb el gen/QTL d'interès, i en pot permetre fins i tot la identificació i la d'allels amb diferents efectes (Thornsberry *et al.*, 2001).

El manteniment del DL depèn de diversos factors relacionats amb la història de l'espècie estudiada (autogàmia o al·logàmia, mida de les poblacions, selecció, etc.). En funció d'aquests paràmetres es poden elaborar col·leccions de genotips amb diferents nivells de DL. Col·leccions amb DL més elevats poden permetre estudis d'associació del genoma sencer, quan es disposa d'un joc de marcadors que cobreixin el genoma a distàncies curtes, fet que pot permetre estudiar qualsevol caràcter que sigui variable en la col·lecció de plantes analitzada, sense la limitació del nombre d'allels (freqüentment només dos) de les poblacions de mapatge convencional. D'altra banda, el genotipatge de la col·lecció serveix per analitzar posteriorment qualsevol caràcter d'interès, qualitatiu o quantitatiu.

Tot i els avantatges de les associacions marcador-caràcter, les anàlisis d'associació tenen algunes dificultats tècniques importants, la més destacable de les quals és l'existència d'estructures subpoblacionals dins de l'espècie, que generen associacions espúries. L'estructura s'ha de tenir en compte, però es pot estudiar coneixent la

variabilitat de cada collecció d'acord amb la variabilitat de marcadors analitzats prèviament, i se'n poden minimitzar els efectes. Un altre element important és que els jocs de marcadors necessaris (previsiblement SNP) ben distribuïts pels cromosomes (el que s'anomena *mapa d'haplotips*) per fer una exploració de tot el genoma són necessàriament molt grans (desenes o fins i tot centenars de milers), cosa que de moment queda fora de l'abast de la majoria d'espècies però que, com molts altres desenvolupaments, sembla possible per a molts conreus a mitjà o llarg termini.

Ús dels marcadors moleculars en la introgressió de gens

Un dels objectius més freqüents en la millora de plantes és la introgressió d'un gen d'un material exòtic o una espècie silvestre dins d'una línia d'alta qualitat o una varietat. Aquesta és la via més freqüent d'utilització de la variabilitat natural per introduir gens interessants dins del genoma d'una espècie cultivada. L'objectiu és complementar un genotip de bon comportament general comprovat amb algun gen que el millori d'una manera puntual com, per exemple, conferint-li resistència a una malaltia o allargant la vida postcollita del seu producte. La línia millorada és en essència una NIL de la primera, que incorpora un fragment de DNA, en aquest cas tan petit com sigui possible, on hi ha el gen que ens interessa. El mètode utilitzat, anomenat *de retroencreuament*, consisteix, d'una manera semblant al que s'ha explicat abans per construir una NIL, en el fet que la línia de bon comportament (R) és hibridada amb la línia exòtica donadora del gen (D) i la línia millorada s'obté després d'una sèrie de retroencreuaments amb R en cadascun dels quals se selecciona a favor del gen introgredit i, al mateix temps, es retorna a les característiques de la línia R, fet

que s'acostuma a completar després de 6-8 generacions de retroencreuament.

La selecció del gen a introgredit es fa quan és possible amb un marcador que estigui a la vora del gen. Els marcadors es poden usar també al mateix temps per seleccionar a favor del genoma sencer dels individus segregants, de manera que sigui possible arribar més de pressa al genotip de la línia R. Tanksley *et al.* (1981), que foren els primers a usar aquesta aproximació, van estimar en un genoma com el del tomàquet que la selecció de només un marcador per cromosoma a favor del genotip del genitor recurrent duria a una fixació del 95 % del genoma en la primera generació de retroencreuament, quan aquesta proporció sense selecció assistida amb marcadors seria només del 75 %. L'ús de marcadors seleccionats de manera que cobreixin el genoma a distàncies de 20-30 cM millora substancialment aquestes previsions, de manera que el procés d'introgressió d'un gen nou es pot reduir amb aquesta aproximació, anomenada *selecció de tot el genoma*, de 6-8 generacions a només dues o tres (Tanksley *et al.*, 1989).

Un altre element important del procés d'introgressió és la mida del fragment introgredit. Stam i Zeven (1981) van estimar que després de vint generacions de retroencreuament encara podia quedar un fragment de D tan gran com 10 cM al voltant del gen seleccionat. Això vol dir que juntament amb el fragment introgredit s'incorpora una gran quantitat de DNA no desitjat que pot incloure centenars de gens de D. Aquesta és la causa de l'associació freqüent de determinats gens introgredits amb característiques indesitjables en altres aspectes del comportament de la línia seleccionada, que s'anomena *linkage drag* (que es podria traduir com 'ròssec de lligament'). Aquest efecte augmenta amb la llunyania entre les línies donant i recurrent, i amb el nombre de gens a introgredit en un sol encreuament.

Els marcadors són útils per a la reducció del ròssec de lligament. Si hi ha dos marcadors lligats i flanquejant el gen d'interès, es pot seleccionar en algun moment del programa de lligament en plantes que tinguin aquest gen però l'hagin recombinat amb un dels dos marcadors (un 1 % de les plantes analitzades si la distància fos d'1 cM), i per a l'altre en la generació següent. D'aquesta manera el fragment introgredit quedarà reduït a menys de la distància entre els dos marcadors flanquejants (Tanksley *et al.*, 1989).

CONCLUSIÓ

La selecció assistida amb marcadors i altres aplicacions dels marcadors en diverses activitats del procés de millora han estat adoptats d'una manera general per part de les empreses que treballen en espècies que es reproduïxen per llavors, principalment els conreus extensius (cereals, oleaginoses, etc.) i hortalisses. En la major part dels casos s'usen per a activitats relacionades amb el control de la qualitat del material que comercialitzen (test de puresa híbrida en híbrids F1, control de la identitat de línies pures i híbrides, defensa de les patents de varietats vegetals), en la selecció de gens majors (freqüentment resistències a malalties) i en la selecció de tot el genoma en programes d'introgressió de gens. També en la selecció de caràcters quantitius, encara que aquesta aplicació, més complexa i amb una inversió inicial en recerca, està limitada a algunes espècies més conegudes des del punt de vista genètic (blat de moro, blat, arròs, ordi, tomàquet o colza, entre altres), normalment en els programes de les empreses més grans. En el cas dels conreus que es reproduïxen vegetativament (fruiters, vinya, maduixa, ornamentals) les empremtes de DNA obtingudes amb marcadors (principalment SSR) s'usen a bastament per a la protecció dels drets d'obtentor, però

per ara molt poc per a la selecció, ja que es coneix poc sobre la genètica d'aquestes espècies, la metodologia de millora emprada habitualment és poc propícia per al seu ús i les empreses que es dediquen a aquestes espècies són habitualment petites i menys tecnificades.

El coneixement del fenotip necessita una mesura acurada, fet que segueix sent un element clau perquè es pugui associar aquesta informació amb la que s'obtingui del genotip, és a dir, en la seqüència del DNA. Si bé el progrés del coneixement del genotip ha estat molt important en els darrers anys, una part molt gran del que puguem fer en el futur dependrà també d'un augment en quantitat i qualitat de les dades actuals sobre variabilitat fenotípica dins del germoplasma cultivat i silvestre pròxim de cada conreu, així com de la creació i caracterització de poblacions segregants de mapatge adequades per a la dissecció de caràcters quantitius, el desenvolupament de col·leccions de mutants i el de poblacions adequades per a la genètica d'associació. La informació sobre el genotip ha estat útil per a una interpretació millor de les dades fenotípiques, tal com ha demostrat l'anàlisi de QTL, i l'avenç metodològic de les eines d'anàlisi de molècules ha de permetre una caracterització a fons del fenotip pel que fa al metaboloma complet. En l'àmbit genòmic el desplegament de les noves tècniques de seqüenciació massiva fa preveure que els propers anys es produirà la seqüenciació (i reseqüenciació) de molts dels genomes de les espècies conreades i augmentarà en gran mesura la comprensió de la funció de molts gens i de les seves interaccions. Tota aquesta informació generarà un major grau d'aplicació en la millora de les eines que ja existeixen, ja que simplificarà la cerca de gens i marcadors per a aquests gens i ampliarà el nombre de caràcters a seleccionar i el d'espècies en què aquests se seleccionaran. En conjunt, el progrés metodològic i

de coneixement del genoma, que ja ha estat enorme en els darrers vint anys, promet seguir a un ritme creixent en el futur immediat, cosa que ens ha de permetre afinar cada vegada més en la predicció del fenotip a partir del coneixement del genotip, amb uns criteris de rapidesa i baix cost, que es troben en la base de l'ús dels marcadors en la millora genètica.

BIBLIOGRAFIA

- AHMADI, N.; ALBAR, L.; PRESSOIR, G.; PINEL, A.; FARGETTE, D.; GHESQUIERE, A. (2001). «Genetic basis and mapping of the resistance to Rice yellow mottle virus. III. Analysis of QTL efficiency in introgressed progenies confirmed the hypothesis of complementary epistasis between two resistance QTLs». *Theor. Appl. Genet.*, 103: 1084-1092.
- AHN, S.; TANKSLEY, S. D. (1993). «Comparative linkage maps of the rice and maize genomes». *Proc. Nat. Acad. Sci. USA*, 90: 7980-7984.
- AKOPYANZ, N.; BUKANOV, N. O.; WESTBLOM, T. U.; BERG, D. E. (1992). «PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*». *Nucleic Acid Res.*, 20: 6221-6225.
- ALDERBORN, A.; KRISTOFFERSON, A.; HAMMERLING, U. (2000). «Determination of single-nucleotide polymorphisms by real-time pyrophosphate DNA sequencing». *Genome Res.*, 10: 1249-1258.
- BARBAZUK, W. B.; EMRICH, S. J.; CHEN, H. D.; LI, L.; SCHNABLE, P. S. (2007). «SNP discovery via 454 transcriptome sequencing». *Plant J.*, 51: 910-918.
- BARLOY, D.; LEMOINE, J.; ABELARD, P.; TANGUY, A. M.; RIVOAL, R.; JAHIER, J. (2007). «Marker-assisted pyramiding of two cereal cyst nematode resistance genes from *Aegilops variabilis* in wheat». *Molec. Breed.*, 20: 31-40.
- BARNES, S. (2002). «Comparing *Arabidopsis* to other flowering plants». *Curr. Opinion Plant Biol.*, 5: 128-133.
- BERNATZKY, R.; TANKSLEY, S. D. (1986). «Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences». *Genetics*, 112: 887-898.
- BILLOTE, N.; LAGODA, P. J. L.; RISTERUCCI, A. M.; BAURENS, F. C. (1999). «Microsatellite-enriched libraries: Applied methodology for the development of SSR markers in tropical crops». *Fruits*, 54: 277-288.
- BOREVITZ, J. O.; LIANG, D.; PLOUFFE, D.; CHANG, H. S.; ZHU, T.; WEIGEL, D.; BERRY, C. C.; WINZELER, E.; CHORY, J. (2003). «Large-scale identification of Single-Feature Polymorphisms in complex genomes». *Genome Res.*, 13: 513-523.
- BOTSTEIN, D.; WHITE, R. L.; SKOLNICK, M.; DAVIS, R. W. (1980). «Construction of a genetic-linkage map in man using restriction fragment length polymorphisms». *Am. J. Hum. Genet.*, 32: 314-331.
- CAI, D.; KLEINE, M.; KIFLE, S.; HARLOFF, H. J.; SANDAL, N. N.; MARCKER, K. A.; KLEIN-LANKHORST, R. M.; SALLENTIJN, E. M. J.; LANGE, W.; STIEKEMA, W. J.; WYSS, U.; GRUNDLER, F. M. W.; JUNG, C. (1997). «Positional cloning of a gene for nematode resistance in sugar beet». *Science*, 275: 832-834.
- CASTRO, A.; CHEN, X.; COREY, A. E.; FILICHKINA, T.; HAYES, P. M.; MUNDT, C.; RICHARDSON, K.; SANDOVAL-ISLAS, S.; VIVAR, H. (2003). «Pyramiding and validation of Quantitative Trait Locus (QTL) alleles determining resistance to barley stripe rust: effects on adult plant resistance». *Crop Sci.*, 43: 2234-2239.
- CHANG, Y. L.; TAO, Q.; SCHEURING, C.; DING, K.; MEKSEM, K.; ZHANG, H. B. (2001). «An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence». *Genetics*, 159: 1231-1242.
- DEVOS, K. M. (2005). «Updating the crop circle». *Plant Biol.*, 8: 155-162.
- ESHED, Y.; ZAMIR, D. (1995). «An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL». *Genetics*, 141: 1147-1162.
- FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. (2003). «Structure of linkage disequilibrium in plants». *Annu. Rev. Plant Biol.*, 54: 357-374.
- FRIDMAN, E.; PLEBAN, T.; ZAMIR, D. (2000). «A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene». *Proc. Natl. Acad. Sci. USA*, 97: 4718-4723.
- GARCIA-MAS, J.; GRAZIANO, E.; ARANZANA, M. J.; MONFORTE, A.; OLIVER, M.; BALLESTER, J.; VIRUEL, M. A.; ARÚS, P. (2000). «Marcadores de ADN: concepto, tipos, protocolos». A: NÚEZ, F.; CARRILLO, J. M. [ed.]. *Los marcadores genéticos en mejora vegetal*. València: UPV: 93-151.
- GOTTLIEB, L. D. (1981). «Electrophoretic evidence and plant populations». *Phytochemistry*, 7: 1-45.
- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. (2005). «Linkage disequilibrium and association studies in higher plants: present status and future prospects». *Plant Molec. Biol.*, 57: 461-485.
- JENSEN, R. C.; NAP, J. P. (2001). «Genetical genomics: the added value from segregation». *Trends Genet.*, 17: 388-391.
- KU, H. M.; LIU, J. P.; DOGANLAR, S.; TANKSLEY, S. D. (2001). «Exploitation of *Arabidopsis*-tomato synteny

- to construct a high-resolution map of the ovate-containing region in tomato chromosome 2». *Genome*, 44: 470-475.
- LEWONTIN, R. C. (1970). *The Genetic Basis of Evolutionary Change*. Nova York: Columbia University Press.
- LIVAK, K. J. (1999). «Allelic discrimination using fluorogenic probes and the 5' nuclease assay». *Genet. Anal. Tech. Appl.*, 14: 143-149.
- MACMILLAN, K.; EMRICH, K.; PIEPHO, H. P.; MULLINS, C. E.; PRICE, A. H. (2006). «Assessing the importance of genotype X environment interaction for root traits in rice using a mapping population II: conventional QTL analysis». *Theor. Appl. Genet.*, 113: 953-964.
- MARTIN, G. B.; BROMMONSCHENKEL, S. H.; CHUNWONGSE, J.; FRARY, A.; GANAL, M. W.; SPIVEY, R.; WU, T. Y.; EARLE, E. D.; TANKSLEY, S. D. (1993). «Map-based cloning of a protein kinase gene conferring disease resistance in tomato». *Science*, 262: 1432-1436.
- MICHELMORE, R. W.; PARAN, I.; KESSELL, R. V. (1991). «Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations». *Proc. Natl. Acad. Sci. USA*, 88: 1212-1217.
- MORGANTE, M.; HANAFEY, M.; POWELL, W. (2002). «Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes». *Nature Genet.*, 30: 194-200.
- MORGANTE, M.; OLIVIERI, A. M. (1993). «PCR-amplified microsatellites as markers in plant genetics». *Plant J.*, 3: 175-182.
- O'BRIEN, S. J. (1990). *Genetic maps: locus maps of complex genomes*. Nova York: Cold Spring Harbor Laboratory Press.
- O'DONOVAN, M. C.; OEFNER, P. J.; ROBERTS, S. C.; AUSTIN, J.; HOOGENDOORN, B.; GUY, C.; SPEIGHT, G.; UPADHYAYA, M.; SOMMER, S. S.; MCGUFFIN, P. (1998). «Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection». *Genomics*, 52: 44-49.
- OLSEN, M.; HOOD, L.; CANTOR, C.; BOTSTEIN, D. (1989). «A common language for physical mapping of the human genome». *Science*, 245: 1434-1435.
- PARAN, I.; MICHELMORE, R. W. (1993). «Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce». *Theor. Appl. Genet.*, 85: 985-993.
- PATERSON, A. H.; LANDER, E. S.; HEWITT, J. D.; PETERSON, S.; LINCOLN, S. E.; TANKSLEY, S. D. (1988). «Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms». *Nature*, 335: 721-726.
- PFLIEGER, S.; LEFEBVRE, V.; CAUSSE, M. (2001). «The candidate gene approach in plant genetics: a review». *Molec. Breed.*, 7: 275-291.
- RODI, C. P.; DARNHOFFER-PATEL, B.; STANSSENS, P.; ZABEAU, M.; BOOM, D. VAN DER (2002). «A strategy for the rapid discovery of disease markers using the MassARRAY system». *BioTechniques*, 32: 62-69.
- SAIKI, R. K.; GELFAND, D. H.; STOFFEL, S.; SCHARF, S. J.; HIGUCHI, R.; HORN, G. T.; MULLIS, K. B.; ERlich, H. A. (1988). «Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase». *Science*, 239: 487-491.
- STAM, P.; ZEVEN, C. (1981). «The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing». *Euphytica*, 30: 227-238.
- SOUTHERN, E. M. (1975). «Detection of specific sequences among DNA fragments separated by gel-electrophoresis». *J. Molec. Biol.*, 98: 503.
- TANKSLEY, S. D. (1983). «Molecular markers in plant breeding». *Plant. Molec. Biol. Rep.*, 1: 3-8.
- (1993). «Mapping polygenes». *Annu. Rev. Genet.*, 27: 205-233.
- TANKSLEY, S. D.; GANAL, M. W.; MARTIN, G. B. (1995). «Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes». *Trends Genet.*, 11: 63-68.
- TANKSLEY, S. D.; MEDINA-FILHO, H.; RICK, C. M. (1981). «The effect of isozyme selection on metric characteristics in an interspecific backcross of tomato - basis of an early screening procedure». *Theor. Appl. Genet.*, 50: 291-296.
- (1982). «Use of naturally-occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato». *Heredity*, 49: 11-25.
- TANKSLEY, S. D.; NELSON, J. C. (1996). «Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines». *Theor. Appl. Genet.*, 92: 59-87.
- TANKSLEY, S. D.; ORTON, T. J. (1983). *Isozymes in Plant Genetics and Breeding. Parts A and B*. Nova York: Elsevier Science.
- TANKSLEY, S. D.; YOUNG, N. D.; PATERSON, A. H.; BONIERBALE, M. W. (1989). «RFLP mapping in plant breeding: new tools for an old science». *BioTechnology*, 7: 257-263.
- TAR'AN, B.; BUCHWALDT, L.; TULLU, A.; BANNIZA, S.; WARKENTIN, T. D.; VANDENBERG, A. (2003). «Using molecular markers to pyramid genes for resistance to ascochyta blight and anthracnose in lentil (*Lens culinaris* Medik)». *Euphytica*, 134: 223-230.
- THORNSBERRY, J. M.; GOODMAN, M. M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E. S. (2001).

- «Dwarf8 polymorphisms associate with variation in flowering time». *Nature Genet.*, 28: 286-289.
- VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; LEE, T. V. D.; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, H.; KUIPER, M.; ZABEAU, M. (1995). «AFLP: a new technique for DNA fingerprinting». *Nucleic Acids Res.*, 2: 4407-4414.
- WILLIAMS, J. G. K.; KUBELIK, A. R.; LIVAK, K. J.; RAFALSKI, J. A.; TINGEY, W. V. (1990). «DNA polymorphisms amplified by arbitrary primers are useful as genetic markers». *Nucleic Acids Res.*, 18: 6531-6535.
- YOGESWARAN, K.; FRARY, A.; YORK, T. L.; AMENTA, A.; LESSER, A. H.; NASRALLAH, J. B.; TANKSLEY, S. D.; NASRALLAH, M. E. (2005). «Comparative genome analyses of *Arabidopsis* spp.: Inferring chromosomal rearrangement events in the evolutionary history of *Arabidopsis thaliana*». *Genome Res.*, 15: 505-515.
- ZHANG, H. B.; WING, R. A. (1997). «Physical mapping of the rice genome with BACs». *Plant Molec. Biol.*, 35: 115-127.