

## CÁC KỸ THUẬT CHỈ THỊ DNA TRONG NGHIÊN CỨU VÀ CHỌN LỌC THỰC VẬT

Nguyễn Đức Thành

Viện Công nghệ sinh học, Viện Hàn lâm KH & CN Việt Nam, nguyenducthanh\_pcg@ibt.ac.vn

**TÓM TẮT:** Từ thập kỷ 80 của thế kỷ trước, các kỹ thuật chỉ thị DNA được bắt đầu và phát triển nhanh chóng trở thành lĩnh vực quan trọng trong sinh học phân tử. Các chỉ thị DNA được sử dụng rộng rãi trong nghiên cứu và chọn lọc. Các kỹ thuật chỉ thị DNA được xây dựng để phát triển các chỉ thị DNA cho nghiên cứu đa dạng di truyền, phát sinh loài, phân loại, đánh dấu và xác định gen; cho chọn lọc nguồn gen và chọn giống nhờ chỉ thị phân tử. Sự phát triển của hàng loạt các chỉ thị DNA khác nhau cùng với các nguyên lý, phương pháp và ứng dụng của chúng dẫn đến việc cần thiết xem xét một cách cẩn thận trong việc lựa chọn kỹ thuật phù hợp cho mục đích nghiên cứu. Không có chỉ thị DNA nào hiện biết có thể đáp ứng đầy đủ tất cả các yêu cầu của nhà nghiên cứu. Phụ thuộc vào vấn đề nghiên cứu, có thể chọn trong các kỹ thuật chỉ thị DNA khác nhau mà mỗi kỹ thuật có thể có một số đặc tính cần thiết. Ở Việt Nam, một số kỹ thuật chỉ thị DNA được bắt đầu sử dụng từ cuối những năm 90 của thế kỷ XX. Tuy nhiên, việc sử dụng còn hạn chế vì mới chủ yếu sử dụng các kỹ thuật như đa hình DNA nhân bản ngẫu nhiên, kỹ thuật các chuỗi lặp lại đơn giản hay tiêu vệ tinh, kỹ thuật đa hình độ dài các đoạn nhân bản chọn lọc trong các nghiên cứu: đa dạng di truyền, lập bản đồ liên kết phân tử và chọn giống nhờ chỉ thị phân tử ở thực vật. Bài tổng quan này sẽ giới thiệu tổng quát về hầu hết các kỹ thuật chỉ thị DNA hiện có và ứng dụng của chúng trong nghiên cứu và chọn lọc ở thực vật nhằm cung cấp thông tin cần thiết và cập nhật cho các nhà nghiên cứu phân loại, bảo tồn và chọn giống trong việc lựa chọn kỹ thuật chỉ thị DNA phù hợp.

*Từ khóa:* Chọn giống nhờ chỉ thị phân tử, chọn lọc nguồn gen, đa dạng di truyền, kỹ thuật chỉ thị DNA, xác định gen.

### MỞ ĐẦU

Kể từ khi chỉ thị DNA đầu tiên được phát triển và ứng dụng cho đến cuối những năm 90 của thế kỷ XX, hàng loạt chỉ thị DNA hay còn gọi là chỉ thị phân tử được ra đời đó là các chỉ thị: chỉ thị đa hình độ dài các đoạn cắt hạn chế (RFLP-restriction fragment length polymorphism [20], chuỗi lặp ngắn liền kề (STR-short tandem repeats [64], số lượng thay đổi các chuỗi lặp lại liền kề (VNTR-variable number tandem repeat [127], chỉ thị nhân bản allen đặc biệt (AS-PCR-allele specific polymerase chain reaction [97], các oligo allen đặc biệt (ASO-allele specific oligo [15], đa hình cấu tạo sợi đơn (SSCP-single stranded conformation polymorphism [136], vị trí chuỗi đánh dấu (STS-sequence tagged site [134], đa hình DNA nhân bản ngẫu nhiên (RAPD-random amplified polymorphic DNA [197], chỉ thị tạo ra do phản ứng PCR với các môi tùy tiện (AP-PCR-arbitrarily primed PCR [195], vị trí trình tự tiêu vệ tinh được đánh dấu (STMS-sequence tagged microsatellite site [16], dấu DNA nhân

bản (DAF-DNA amplification fingerprinting [28], chỉ thị trình tự biểu hiện (EST-expressed sequence tags [2], đa hình độ dài các chuỗi đơn giản (SSLP-simple sequence length polymorphism [38], chuỗi đa hình nhân bản được cắt hạn chế (CAPS-cleaved amplified polymorphic sequence [6], chỉ thị tạo ra do phản ứng PCR với môi thoái hóa-môi có vị trí mà ở đó có thể thay thế các oligo khác nhau (DOP-PCR-degenerate oligonucleotide primed PCR [170], chỉ thị nhân bản sợi thay thế (SDA-strand displacement amplification [190], chuỗi lặp lại đơn giản (SSR-simple sequence repeat [5], vùng nhân bản chuỗi DNA được mô tả (SCAR-sequence characterized amplified regions [137], đa hình nucleotide đơn (SNP-single nucleotide polymorphism [78], chỉ thị phản ứng nhân bản bằng môi đơn (single primer amplification reactions [60], đa hình các locus tiêu vệ tinh nhân bản chọn lọc (SAMPL-selective amplification of microsatellite polymorphic loci [122], tiêu vệ tinh neo nhân bản (AMP-PCR-anchored microsatellite primed PCR [212], đa

hình tiêu vệ tinh nhân bản ngẫu nhiên (RAMP-random amplified microsatellite polymorphism [201], chuỗi lặp lại đơn giản giữa (ISSR-inter-simple sequence repeat [212]), các môi liên quan đến allen đặc biệt (ASAP-allele specific associated primers [57], đa hình độ dài đoạn nhân chọn lọc (AFLP-amplified fragment length polymorphism [200], chỉ thị PCR lồng vị trí chọn lọc (SSI-site-selected insertion PCR [92], tiêu vệ tinh nhân bản ngẫu nhiên (RAM-random amplified microsatellites [66], đa hình nhân bản chuỗi đặc biệt (S-SAP-sequence specific amplification polymorphism [193], các trình tự lặp lại đơn giản neo (ASSR-anchored simple sequence repeats [191] và chỉ thị PCR đảo (IPCR-inverse PCR [187]. Các kỹ thuật chỉ thị DNA tương ứng được xây dựng để phát triển chỉ thị DNA cho nghiên cứu đa dạng di truyền, phát sinh loài, phân loại, đánh dấu và xác định gen; cho chọn lọc nguồn gen và chọn giống nhờ chỉ thị phân tử. Không có chỉ thị DNA nào hiện có có thể đáp ứng đầy đủ tất cả các yêu cầu của nhà nghiên cứu. Phụ thuộc vào vấn đề nghiên cứu, ta có thể chọn trong các kỹ thuật chỉ thị DNA khác nhau mà mỗi kỹ thuật có thể có một số đặc tính cần thiết. Ở Việt Nam, một số kỹ thuật chỉ thị DNA được bắt đầu sử dụng từ cuối những năm 90 của thế kỷ XX. Tuy nhiên, việc sử dụng còn hạn chế vì mới chủ yếu sử dụng các kỹ thuật như đa hình DNA nhân bản ngẫu nhiên, kỹ thuật các chuỗi lặp lại đơn giản hay tiêu vệ tinh, kỹ thuật đa hình độ dài các đoạn nhân bản chọn lọc trong các nghiên cứu: đa dạng di truyền và đánh giá nguồn gen [22, 37, 171, 172], lập bản đồ liên kết phân tử [98, 173] và chọn giống nhờ chỉ thị phân tử ở thực vật [99, 104]. Bài tổng quan này sẽ giới thiệu tổng quát về hầu hết các kỹ thuật chỉ thị DNA hiện có và ứng dụng của chúng trong nghiên cứu và chọn lọc ở thực vật nhằm cung cấp thông tin cần thiết và cập nhật cho các nhà nghiên cứu và chọn giống trong việc lựa chọn kỹ thuật chỉ thị DNA phù hợp cho nghiên cứu và chọn lọc ở thực vật.

### **Chỉ thị di truyền, chỉ thị phân tử và chỉ thị DNA**

Các chỉ thị di truyền là các chỉ thị thể hiện sự khác biệt giữa các cơ thể hoặc các loài khác nhau. Các chỉ thị này không thể hiện cho các

gen mục tiêu mà chỉ là dấu hiệu hoặc như là “cờ đánh dấu”. Chỉ thị di truyền bao gồm cả chỉ thị hình thái và chỉ thị phân tử (isozyme, protein, DNA). Tất cả chỉ thị di truyền đều chiếm một vị trí đặc biệt trong nhiễm sắc thể (NST) và được gọi là locus. Các chỉ thị di truyền thường liên kết với gen và được di truyền theo qui luật di truyền. Chỉ thị phân tử thường được hiểu là chỉ thị DNA, các chỉ thị này chỉ nằm gần hay liên kết với gen và không có hoặc ít ảnh hưởng đến kiểu hình.

Chỉ thị di truyền có thể chia thành hai loại: chỉ thị truyền thống và chỉ thị DNA. Chỉ thị truyền thống gồm chỉ thị hình thái (đây là những tính trạng hoặc đặc điểm hình thái như màu hoa, kiểu hình hạt, đặc điểm sinh trưởng, sự biến đổi sắc tố v.v.), chỉ thị tế bào (đặc trưng cấu trúc của nhiễm sắc thể) và chỉ thị sinh hóa (các isozyme, protein và các chất trao đổi chất). Chỉ thị DNA là những thay đổi trong phân tử DNA và được chia thành nhiều loại dựa trên sự khác nhau về phương pháp và kỹ thuật xác định sự đa hình.

Các chỉ thị DNA được sử dụng rộng rãi nhất do số lượng chỉ thị không hạn chế. Chỉ thị DNA được hình thành từ các loại đột biến DNA khác nhau như thay thế (đột biến điểm), sắp xếp lại (thêm vào hoặc bớt đi nucleotide) hoặc các sai sót trong sao chép các đoạn DNA lặp lại liên kề. Các chỉ thị DNA thường nằm ở các vùng không phiên mã. Khác với các chỉ thị hình thái và sinh hóa, chỉ thị DNA không giới hạn về số lượng, không ảnh hưởng bởi yếu tố môi trường và giai đoạn phát triển của cây. Chỉ thị DNA được sử dụng nhiều trong nghiên cứu quan hệ di truyền, phát sinh chủng loại và phân loại phân tử; trong lập bản đồ liên kết di truyền, nhận biết gen; và trong chọn giống bao gồm đánh giá đa dạng di truyền, nhận biết giống, chọn lọc các tính trạng kháng bệnh, chống chịu các điều kiện bất lợi của môi trường, năng suất và phẩm chất giống.

### **Các kỹ thuật chỉ thị DNA**

Mỗi loại chỉ thị DNA được phát triển bằng một kỹ thuật tương ứng. Kỹ thuật chỉ thị DNA lý tưởng cần phải có các tiêu chí sau: cho đa hình cao và phân bố đều trong genome; cho sự phân biệt rõ sự khác nhau về di truyền, tạo nhiều chỉ thị độc lập và chính xác; đơn giản, nhanh và ít

tồn kém; cần ít mẫu và DNA; liên kết với kiểu hình nhất định; có thể lặp lại trong các nghiên cứu, mức độ sai sót thấp nhất, ghi số liệu dễ và chính xác, có nhiều allen (hàm lượng thông tin cao), không cần biết trước thông tin về genome

và cơ thể. Trong bảng 1 là các các kỹ thuật chỉ thị DNA hiện có. Trong bài này tác giả chỉ giới thiệu khái quát về cơ sở, nguyên lý, một số bước cơ bản, những ưu nhược điểm và ứng dụng của một số kỹ thuật phổ biến.

Bảng 1. Các kỹ thuật chỉ thị DNA

Ký hiệu	Tên tiếng Anh	Tên tiếng Việt
<i>Kỹ thuật không sử dụng PCR</i>		
RFLP	Restriction Fragment Length Polymorphism	Đa hình độ dài đoạn cắt hạn chế
REF	Restriction Endonuclease Fingerprinting	Lấy dấu cắt hạn chế
<i>Kỹ thuật sử dụng PCR</i>		
AFLP	Amplified Fragment Length Polymorphism	Đa hình độ dài nhân bản chọn lọc
RAPD	Randomly Amplified Polymorphic DNA	DNA đa hình được nhân bản ngẫu nhiên
AP-PCR	Arbitrarily primed PCR	PCR với mỗi ngẫu nhiên
ARMS	Amplification Refractory Mutation System	Hệ thống đột biến chịu nhiệt nhân bản
ASAP	Arbitrary Signatures from Amplification	Các dấu hiệu ngẫu nhiên từ nhân bản
SPAR	Single Primer Amplification Reaction	Phản ứng nhân bản với mỗi đơn
SPLAT	Single Polymorphic Amplification Test	Phép thử nhân bản đa hình đơn
S-SAP	Sequence-Specific Amplification Polymorphisms	Đa hình nhân bản chuỗi đặc trưng
ASLP	Amplified Sequence Length Polymorphism	Đa hình độ dài chuỗi nhân bản
CAPS	Cleaved Amplification Polymorphic Sequence	Chuỗi đa hình nhân bản được cắt hạn chế
CAS	Coupled Amplification and Sequencing	Giải trình tự và nhân bản kết hợp
DAF	DNA Amplification Fingerprint	Dấu nhân bản DNA
SAMPL	Selective Amplification of Polymorphic Loci	Nhân bản chọn lọc các locus đa hình
<i>Kỹ thuật chỉ thị dựa trên các tiểu vệ tinh</i>		
ISSR	Inter-Simple Sequence Repeats	Chuỗi lặp lại đơn giản giữa
ISTR	Inverse Sequence-Tagged Repeats	Chuỗi lặp lại ngược được đánh dấu
MP-PCR	Microsatellite-Primed PCR	PCR với mỗi tiểu vệ tinh
RAHM	Randomly Amplified Hybridizing Microsatellites	Tiểu vệ tinh lai được nhân bản ngẫu nhiên
RAMPs	Randomly Amplified Microsatellite Polymorphisms	Đa hình tiểu vệ tinh được nhân bản ngẫu nhiên
VNTR	Variable Number Tandem Repeats	Số lượng thay đổi các chuỗi lặp lại liên kế
<i>Các kỹ thuật chỉ thị PCR các chuỗi đặc trưng</i>		
STS	Sequence-Tagged-Site	Vị trí chuỗi đánh dấu
SCAR	Sequence Characterised Amplification Regions	Vùng nhân bản chuỗi được mô tả
SSLP	Single Sequence Length Polymorphism	Đa hình độ dài chuỗi đơn giản
SSR	Simple Sequence Repeats	Các chuỗi lặp lại đơn giản
STMS	Sequence-Tagged Microsatellite Site	Vị trí chuỗi tiểu vệ tinh đánh dấu
<i>Các kỹ thuật chỉ thị gen nhảy</i>		
REMAP	Retrotransposon-Microsatellite Amplified Polymorphism	Đa hình tiểu vệ tinh gen nhảy ngược được nhân bản
RBIP	Retrotransposon-Based Insertion Polymorphism	Đa hình sự gắn dựa trên gen nhảy ngược
IRAP	Inter- Retrotransposon Amplified Polymorphism	Đa hình gen nhảy ngược giữa được nhân bản
TD	Transposable Display	Biểu lộ gen nhảy
<i>Các kỹ thuật chỉ thị nhân khác</i>		
ITS	Internal Transcribed Spacer	Vùng đệm trong được sao mã
SNP	Single Nucleotide Polymorphism	Đa hình nucleotide đơn
OLA	Oligonucleotide Ligation Assay	Phân tích gắn Oligonucleotide
SSCP	Single Stranded Conformation Polymorphism	Đa hình cấu tạo sợi đơn
ASO	Allele Specific Oligonucleotide	Oligonucleotide đặc trưng allen
ASH	Allele-Specific Hybridization	Lai allen đặc trưng

Các kỹ thuật chỉ thị lục lạp		
cpSSR	Chloroplast Simple Sequence Repeats	Chuỗi lặp lại đơn giản lục lạp
RFLPA	Restriction Fragment Length Polymorphism	Phân tích đa hình độ dài đoạn cắt giới hạn DNA lục lạp
cpDNA	Analysis cpDNA	Các RNA vận chuyển lục lạp
tRNAs	Chloroplast Transfer RNAs	
Công nghệ hỗ trợ hiện đại		
DarT	Diversity array Technology	Công nghệ sắp xếp đa dạng
NGS	Next-generation sequencing	Giải trình tự thế hệ thứ hai

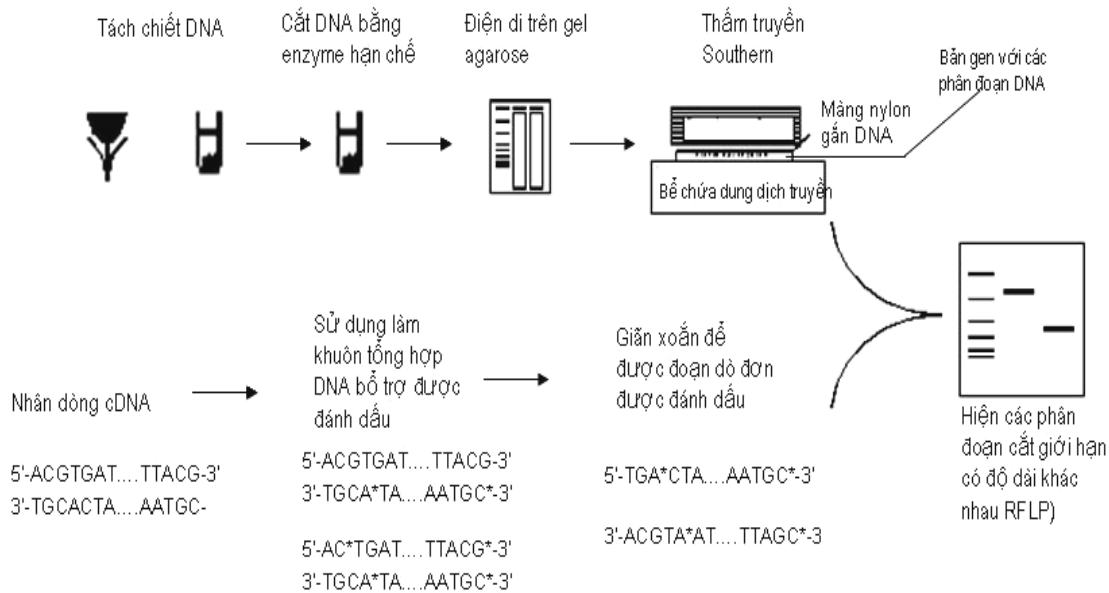
Các kỹ thuật chỉ thị DNA được chia thành các nhóm chính là: các kỹ thuật không sử dụng phản ứng PCR và các kỹ thuật sử dụng PCR. Ngoài ra, cần phải kể đến một số công nghệ hỗ trợ hết sức hiệu quả cho phát triển chỉ thị và nhận biết đa hình chỉ thị DNA như công nghệ sắp xếp đa dạng (Diversity array Technology) và công nghệ giải trình tự thế hệ thứ hai (Next-generation sequencing).

**Các kỹ thuật chỉ thị DNA không sử dụng PCR**

**Kỹ thuật đa hình độ dài đoạn cắt hạn chế (RFLP)**

Trong kỹ thuật đa hình độ dài đoạn cắt hạn chế (Restriction Fragment Length Polymorphism-RFLP), đa hình DNA được xác định bằng cách lai đoạn dò DNA (DNA probe)

đánh dấu với DNA sau khi cắt hạn chế bằng enzyme cắt hạn chế và được thẩm truyền lên màng lai bằng phương pháp Southern. Kết quả là tạo ra hình ảnh các phân đoạn DNA khác nhau. Các hình ảnh phân đoạn DNA khác nhau này được tạo nên do sự thay thế, thêm vào hay bớt đi của các nucleotide hoặc do đa hình nucleotide đơn. Kỹ thuật RFLP được tiến hành theo các bước: cắt DNA bằng một hoặc vài enzyme cắt hạn chế; các phân đoạn DNA sau đó được phân tách trên gel agarose và thẩm truyền lên màng lai. Việc xác định các phân đoạn DNA được tiến hành bằng lai các phân đoạn DNA này với đoạn dò được đánh dấu huỳnh quang hoặc phóng xạ để có thể phát hiện bằng phản ứng huỳnh quang hoặc phim chụp phóng xạ (hình 1).



**Hình 1.** Các bước tiến hành kỹ thuật RFLP: Sau khi tách chiết, DNA genome được cắt bằng enzyme cắt hạn chế, các phân đoạn DNA sau đó được phân tách bằng điện di trên gel agarose, rồi được thẩm truyền bằng phương pháp Southern lên màng nylon và được lai với các đoạn dò được đánh dấu bằng phóng xạ hoặc huỳnh quang, cuối cùng là hiện trên phim hoặc nhuộm bằng chất hiện màu

Các chỉ thị RFLP có mức đa hình cao, đồng trội (nên có thể phân biệt được các cá thể dị hợp tử và đồng hợp tử) và khả năng lặp lại cao. Kỹ thuật RFLP cũng cho phép phân tích đồng thời nhiều mẫu. Tuy nhiên kỹ thuật này không được dùng rộng rãi vì một số hạn chế như: cần nhiều DNA chất lượng cao, phải phát triển thư viện đoạn dò cho từng loài, cần thông tin về trình tự để tạo đoạn dò, không thuận tiện cho việc tự động hóa, mức độ đa hình và số lượng locus trên lần phân tích thấp, đòi hỏi nhiều thời gian, tốn kém. Trong những thập niên 80 và 90 của thế kỷ trước, kỹ thuật RFLP được sử dụng trong lập bản đồ genome [54, 96], xác định giống [82]. RFLP được dùng trong nghiên cứu quan hệ giữa các nhóm phân loại gần [119], đa dạng di truyền [42], trong lai tạo và chuyển gen vào cá thể khác (introgression) bao gồm cả trôi gen giữa cây trồng và cỏ dại [36].

#### **Các kỹ thuật dựa trên phản ứng PCR**

Với sự phát triển kỹ thuật phản ứng chuỗi trùng hợp (polymerase chain reaction-PCR) đã dẫn đến những tiến bộ vượt bậc trong việc cải tiến các kỹ thuật xây dựng các chỉ thị DNA dựa trên phản ứng PCR [123]. Các kỹ thuật chỉ thị DNA dựa trên phản ứng PCR, ngoài hai kỹ thuật sử dụng rộng rãi là RAPD và AFLP, dựa vào sự nhân bản các chuỗi hoặc vùng DNA khác nhau có thể chia ra một số nhóm như: kỹ thuật chỉ thị các chuỗi đặc trưng, kỹ thuật chỉ thị tiêu vệ tinh, kỹ thuật chỉ thị gen nhảy.

#### **Kỹ thuật đa hình DNA nhân ngẫu nhiên (RAPD)**

Cơ sở của kỹ thuật RAPD là sự nhân bản DNA genome bằng phản ứng PCR với các môi ngẫu nhiên để tạo ra sự đa hình DNA do sự tái sắp xếp hoặc mất nucleotide ở vị trí bắt môi [197].

Mỗi sử dụng cho kỹ thuật RAPD là các môi ngẫu nhiên, thường là 10 nucleotide và có nhiệt độ kéo dài môi thấp (34-37°C). Mặc dù trình tự môi RAPD là ngẫu nhiên nhưng phải đạt được hai tiêu chí là: tỷ lệ GC tối thiểu phải là 40% (thường là 50-80%) và không có trình tự bazơ đầu xuôi và ngược giống nhau. Sản phẩm PCR-RAPD thường được phân tách trên gel agarose 1,5-2,0%.

Kỹ thuật RAPD không cần thông tin về genome của đối tượng nghiên cứu và có thể ứng

dụng cho các loài khác nhau với các môi chung. Hơn nữa, kỹ thuật RAPD đơn giản và dễ thực hiện. Nhưng kỹ thuật RAPD có hạn chế là sản phẩm PCR không ổn định do môi ngắn, nhiệt độ bắt môi thấp; ngoài ra, kỹ thuật này tạo ra các chỉ thị trội do đó không phân biệt được các cá thể dị hợp tử với các cá thể đồng hợp tử. RAPD sử dụng nhiều trong nghiên cứu đa dạng di truyền giữa các loài thực vật [84, 85, 176, 177, 179, 180, 185], trong nghiên cứu đặc điểm của giống và đánh giá biến đổi di truyền [114], trong xác định loài [52, 153] và xác định con lai [10, 152].

Kỹ thuật PCR với các môi ngẫu nhiên (Arbitarily Primed PCR-AP-PCR) và lấy dấu bằng nhân bản DNA (DNA amplification fingerprinting-DAF) là hai kỹ thuật được phát triển độc lập và là hai dạng biến thể của kỹ thuật RAPD. Đối với AP-PCR chỉ dùng một môi đơn có độ dài 10 đến 15 nucleotide. Kỹ thuật được thực hiện khác với PCR bình thường là hai chu kỳ đầu được tiến hành ở điều kiện không chặt chẽ (nhiệt độ bắt môi thấp) sau đó ở các chu kỳ sau PCR được tiến hành trong điều kiện chặt chẽ bằng việc tăng nhiệt độ bắt môi. Trong trường hợp DAF thì chỉ sử dụng một môi ngẫu nhiên ngắn hơn 10 nucleotide và sản phẩm PCR được phân tích bằng gel polyacrylamide. Với DAF, mỗi sử dụng ngắn nhưng nồng độ môi cần cao, phản ứng PCR gồm hai chu kỳ nhiệt chứ không phải ba chu kỳ như RAPD. Kỹ thuật AP-PCR khác với RAPD và DAF là: phản ứng PCR chia thành ba bước, mỗi bước có độ chặt chẽ và nồng độ của các thành phần phản ứng khác nhau. Nồng độ môi ở những chu kỳ đầu cao, môi dài 20 nucleotide hoặc hơn.

#### **Kỹ thuật đa hình độ dài đoạn nhân chọn lọc (Amplified Fragment Length Polymorphism-AFLP)**

Kỹ thuật AFLP được sử dụng để phát hiện đa hình DNA. Phân tích AFLP được kết hợp cả RFLP và PCR bằng việc gắn các chuỗi nhận biết vào môi hay còn gọi là chuỗi tiếp hợp (adapter) để nhân chọn lọc các phân đoạn DNA được cắt hạn chế. Các cặp môi thường tạo được từ 50 đến 100 băng trong một phân tích. Số lượng băng phụ thuộc vào số nucleotide chọn lọc trong tổ hợp môi. Kỹ thuật AFLP cho phép lấy dấu DNA từ bất kỳ nguồn gốc nào.

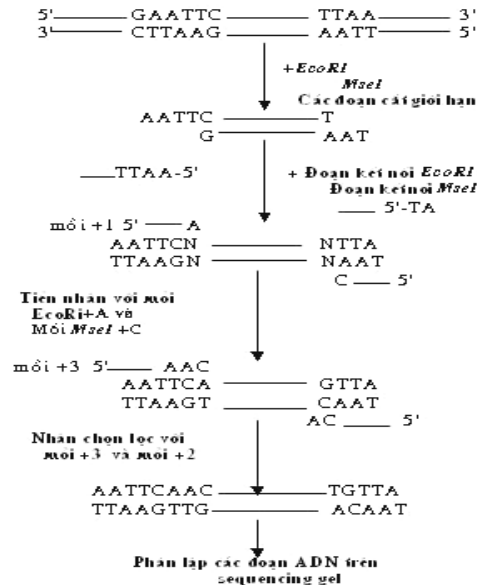
Kỹ thuật AFLP được tiến hành theo một số bước như trong hình 2: đầu tiên DNA genome được cắt bởi enzyme cắt hạn chế cắt không thường xuyên (*EcoRI* hoặc *PstI*) và cắt thường xuyên (*MseI* hoặc *TaqI*). Các phân đoạn tạo ra sau khi cắt hạn chế được gắn với các đoạn kết nối ở cả hai đầu. Các đoạn kết hợp này được thiết kế sao cho vị trí nhận biết ban đầu của enzyme cắt hạn chế không khôi phục lại sau khi gắn. Phản ứng PCR chỉ xảy ra khi các mồi có thể gắn với các phân đoạn có trình tự kết nối cùng với các cặp bazơ bổ sung với các nucleotide chọn lọc. Việc nhân bản được thực hiện hai lần trong điều kiện chặt chẽ với các mồi hỗ trợ với các đoạn tiếp hợp và 1 đến 3 nucleotide chọn lọc ở đầu 3'. Phản ứng PCR lần đầu (tiền nhân) được tiến hành với tổ hợp mồi chứa một nucleotide chọn lọc. Phản ứng PCR lần hai (nhân chọn lọc) được tiến hành với các cặp mồi có 1 đến 3 nucleotide chọn lọc. Do mức độ chọn lọc cao, các mồi khác nhau chỉ một nucleotide sẽ cho các bộ phân đoạn nhân bản khác nhau. Các mồi với 1, 2 và 3 nucleotide chọn lọc sẽ giảm số phân đoạn được nhân bản tương ứng là 4, 16 và 64. Số nucleotide chọn lọc phù hợp thay đổi theo kích thước genome của loài. Các phân đoạn AFLP được phân tách bằng gel polyacrylamide kết hợp nhuộm bạc (AgNO<sub>3</sub>) hoặc bằng máy giải trình tự tự động.

Phân tích AFLP không dễ như RAPD nhưng hiệu quả hơn RFLP. Kỹ thuật này có một số ưu điểm như: có độ tin cậy và lặp lại cao; không cần thông tin về trình tự DNA của cơ thể nghiên cứu; cho nhiều thông tin do có khả năng phân tích số lượng lớn locus đa hình với một tổ hợp mồi trên một gel và có thể cho thấy các locus đặc thù; các số liệu có thể lưu giữ trong cơ sở dữ liệu để so sánh.

Bên cạnh các ưu điểm, AFLP có một số nhược điểm như: phải qua nhiều bước mới có kết quả; DNA cần phải sạch, không có các chất ức chế hoạt động của enzyme cắt hạn chế; đòi hỏi công sức và giá thành cao. Kỹ thuật tạo ra chỉ thị trội, không phân biệt được đồng hợp tử và dị hợp tử.

AFLP được sử dụng trong nghiên cứu lập bản đồ genome [14, 173], đa dạng di truyền [53, 103, 176, 178] và quan hệ chủng loại giữa các

kiểu gen có mối quan hệ gần [69, 159], nghiên cứu cấu trúc di truyền nguồn gen [8, 68] và đánh giá phân hóa di truyền trong quần thể [186, 196].



Hình 2. Các bước tiến hành kỹ thuật AFLP: DNA genome được cắt bằng enzyme *EcoRI* và *MseI* để tạo ra các đoạn cắt hạn chế, các đoạn cắt được gắn với đoạn kết nối *EcoRI* và *MseI*, tiếp theo các đoạn cắt giới hạn có gắn đoạn kết nối được nhân bản bằng PCR với các mồi *EcoRI* và *MseI* tương ứng có thêm một nucleotide (tiền nhân), sản phẩm PCR nhận được được nhân với tổ hợp các mồi *EcoRI* và *MseI* có thêm 1 đến 3 nucleotide (nhân chọn lọc), sản phẩm PCR chọn lọc được phân tách trên gel polyacrylamide và nhuộm bạc để quan sát các băng.

Kỹ thuật chỉ thị dựa trên các tiểu vệ tinh (microsatellite) hay chuỗi lặp lại đơn giản (simple sequence repeats).

Đây là nhóm kỹ thuật sử dụng các cặp mồi đặc trưng để nhân các vùng genome có các nucleotides lặp lại ở các vị trí khác nhau bao gồm: kỹ thuật chuỗi lặp lại đơn giản hay còn gọi là kỹ thuật đa hình độ dài các chuỗi đơn giản (Single Sequence Length Polymorphism-SSLP), kỹ thuật vị trí chuỗi tiểu vệ tinh đánh dấu (Sequence-Tagged Microsatellite Site-STMS), kỹ thuật chuỗi lặp lại giữa (Inter-Simple

Sequence Repeat-ISSR) và kỹ thuật đa hình các tiểu vệ tinh được nhân bản ngẫu nhiên (Randomly Amplified Microsatellite Polymorphisms-RAMP). Chúng ta sẽ xem xét lần lượt các kỹ thuật này trong các phần sau.

Kỹ thuật chuỗi lặp lại đơn giản (Simple Sequence Repeats-SSR)

Chuỗi lặp lại đơn giản (Simple Sequence Repeats-SSR) [169], tiểu vệ tinh (microsatellite) [107], chuỗi lặp lại trước sau ngắn (Short Tandem Repeats-STRs) hay đa hình độ dài chuỗi đơn giản (simple sequence length polymorphisms-SSLPs) [115] là sự lặp lại các chuỗi nucleotide ngắn từ 2-6 nucleotide [30].

SSR (SSLP, STMS) trở thành kỹ thuật chỉ thị phân tử quan trọng trong cả động vật và thực vật. SSR rất đa hình do đột biến tác động lên số đơn vị lặp lại. Sự thay đổi hay sự đa hình của SSR là kết quả của sự khác nhau về độ dài các đoạn lặp lại trong genome do quá trình trao đổi chéo không cân hoặc do sự giảm nucleotide trong quá trình sao chép. SSR không những phổ biến mà còn biến động mạnh về số lượng kiểu lặp lại trong genome sinh vật nhân thực. Sự khác nhau allele của SSR là kết quả của sự thay đổi số lượng đơn vị lặp lại trong cấu trúc tiểu vệ tinh. Các chuỗi lặp lại thường đơn giản và cấu tạo bởi 2, 3 hoặc 4 nucleotide.

Kỹ thuật SSR được thực hiện bằng phản ứng PCR với mỗi SSR xuôi và ngược. Sản phẩm PCR được phân tách trên gel polyacrylamide kết hợp nhuộm bạc (AgNO<sub>3</sub>) hoặc bằng máy giải trình tự tự động.

Việc phát triển chỉ thị SSR được tiến hành theo một số bước như: xây dựng thư viện SSR, xác định locus SSR, xác định vùng phù hợp để thiết kế mồi, PCR với các mồi được thiết kế, đánh giá và phân tích mẫu băng, đánh giá đa hình của sản phẩm PCR.

Kỹ thuật SSR có một số ưu việt hơn các chỉ thị khác như: i. Cho nhiều allele trong một locus; ii. Phân bố đều trong genome; iii. SSR cho thông tin cụ thể hơn so với di truyền ty thể theo đường mẹ (vì có mức đột biến cao) và di truyền theo cả bố và mẹ; iv. Là chỉ thị đồng trội; v. Có tính đa hình và đặc thù cao; vi. Có thể lặp lại ở các thí nghiệm, sử dụng ít DNA, rẻ và dễ tiến

hành, có thể phân tích bán tự động, không sử dụng phóng xạ, có thể sử dụng các DNA cổ (ancient DNA-aDNA). SSR có thể phân biệt các cá thể có mối quan hệ gần. Điểm hạn chế quan trọng của kỹ thuật chỉ thị SSR là cần phải đọc trình tự genome để dựa vào đó có thể thiết kế các cặp mồi đặc thù và tối ưu hóa điều kiện các mồi cho từng loài trước khi sử dụng. Hiện nay, SSR là chỉ thị được chọn cho các nghiên cứu hồ sơ pháp lý, di truyền quần thể và nghiên cứu động vật hoang dã. Ở thực vật SSR được sử dụng trong nghiên cứu đa dạng di truyền [53, 71, 109, 117, 121, 171, 172, 175, 183, 210,], trong chọn cặp lai (43, 71), trong xác định con lai [1] và trong lập bản đồ liên kết phân tử [39, 110, 173].

Kỹ thuật đa hình độ dài các chuỗi đơn giản (Single Sequence Length Polymorphism-SSLP) là kỹ thuật chỉ thị tạo các chuỗi lặp lại với độ dài khác nhau nằm giữa hai gen. Các chuỗi có độ dài khác nhau là do số lượng trình tự lặp lại trong chuỗi khác nhau do quá trình trao đổi chéo không cân hoặc do sự giảm nucleotide trong quá trình sao chép. Cả hai quá trình này đều có thể dẫn đến giảm hoặc tăng số lượng đơn vị lặp lại trong genome. Sự khác nhau về độ dài của SSLP được sử dụng để đánh giá sự thay đổi di truyền giữa các cá thể trong loài.

Kỹ thuật SSLP giống như kỹ thuật SSR, cũng được tiến hành bằng các cặp mồi đặc hiệu nhân bản các chuỗi đơn giản nằm giữa hai gen, sản phẩm PCR được phân tách trên gel polyacrylamide.

SSLP đã được sử dụng trong nghiên cứu đa dạng di truyền [12], xác định giống [67], xác định dị hợp tử [135] và lập bản đồ di truyền [5, 202].

Kỹ thuật vị trí chuỗi tiểu vệ tinh đánh dấu (Sequence-Tagged Microsatellite Site-STMS) [21, 181, 182] là kỹ thuật chỉ thị nhân bản các chuỗi lặp lại ở vị trí được đánh dấu bằng các cặp mồi đặc hiệu. Kỹ thuật này cũng được sử dụng nhiều trong nghiên cứu đa dạng di truyền, phân tích genome [86], lập bản đồ di truyền [51, 198] và nhận biết giống [47].

Ưu điểm nổi bật của SSR, SSLP và STMS là các chỉ thị này có thể sử dụng chung cho một số loài cây. Chẳng hạn STMS của cây đậu đồng

(*Pisum sativum* L.) có thể dùng cho cây đậu châu Á (*Cicer arietinum* L.) và STMS của đậu châu Á có thể dùng cho đậu lăng (*Lens culinaris* L.) và đậu tằm thuộc chi *Vicia* [7], STMS của cây họ đậu có thể dùng cho lạc [22].

**Kỹ thuật chuỗi lặp lại đơn giản giữa (Inter-Simple Sequence Repeat-ISSR)**

ISSR là những chỉ thị được nhân bản bằng PCR với một môi hỗ trợ với tiểu vệ tinh (microsatellite) đích. Mỗi băng tương ứng với chuỗi DNA được giới hạn bởi các tiểu vệ tinh đảo ngược. Kết quả dẫn đến đa locus, có sự đa hình cao và tạo ra các chỉ thị trội. ISSR-PCR là kỹ thuật đánh giá kiểu gen nhanh, không đắt; sự đa hình dựa trên sự thay đổi trong các vùng nằm giữa các tiểu vệ tinh. Kỹ thuật này không cần thông tin về trình tự, tạo được nhiều locus, có tính đa hình cao và tạo ra chỉ thị trội [120]. Kỹ thuật ISSR là kỹ thuật nhân bản đoạn DNA nằm giữa hai vùng lặp lại giống hệt và ngược chiều nhau (hình 3). Kỹ thuật này sử dụng các tiểu vệ tinh như các môi trong phản ứng PCR với một môi cho nhiều locus đích để nhân bản chủ yếu các chuỗi lặp lại đơn giản với độ dài khác nhau. Các tiểu vệ tinh sử dụng như môi trong kỹ thuật ISSR có thể là 2, 3, 4, hoặc 5 nucleotide. Các môi sử dụng có thể không phải môi neo hoặc là môi neo ở đầu 3' hoặc 5' với 1 đến 4 nucleotide thoái hóa kéo đến các chuỗi bên cạnh. Kỹ thuật ISSR sử dụng môi dài (15 đến 30 nucleotide) vì vậy nhiệt độ bắt môi cao dẫn đến độ ổn định cao của phản ứng. Sản phẩm có độ dài 200-2000 bp nên có thể phân tách trên cả gel agarose và polyacrylamide.

Kỹ thuật ISSR được sử dụng rất rộng rãi trong nghiên cứu đa dạng di truyền [58, 149], nghiên cứu đặc điểm di truyền trong quần thể [145], lấy dấu di truyền [18, 29], đánh dấu gen [147], xác định cây trồng [47, 108], phân tích nguồn gốc [48], xác định sự thay đổi genome [101] và đánh giá con lai [105]. Kỹ thuật ISSR có một số lợi thế so với các kỹ thuật khác là có thể phân biệt được các kiểu gen gần và không cần thông tin về trình tự gen của cây nghiên cứu. Giống như RAPD, ISSR là kỹ thuật nhanh, dễ tiến hành. Nó ưu việt hơn RAPD là cho nhiều thông tin và có thể lặp lại ở các thí nghiệm do môi dài hơn. Nhược điểm chủ yếu

của ISSR là tạo ra các chỉ thị trội [60, 191] và sự dị đồng nhất của các sản phẩm nhân bản do sự di chuyển đồng thời.



**Hình 3.** Nhân bản chuỗi lặp lại nằm giữa hai chuỗi (CT)<sub>n</sub> ngược chiều nhau bằng môi đơn (GA)<sub>n</sub>: A. Môi đơn (GA)<sub>n</sub> không neo có thể bắt cặp bất kỳ ở vị trí nào trong vùng (CT)<sub>n</sub> của DNA khuôn nên việc nhân bản không được chuẩn, B. Môi đơn (GA)<sub>n</sub> có hai nucleotide (NN) neo ở đầu 3' bắt cặp vào vùng đặc thù trên DNA khuôn và tạo băng rõ ràng, C. (GA)<sub>n</sub> có hai nucleotide (NN) neo ở đầu 5' bắt cặp vào vùng đặc thù trên DNA khuôn nhân bản được một phần vùng lặp lại tạo ra các băng lớn.

ISSR được dùng trong nghiên cứu đa dạng di truyền nhiều loài cây trồng khác nhau như lúa [79, 149], lúa mì [124], kê [156], nho [3], khoai lang [70], mã đề [199], táo ta [131] v.v.

**Kỹ thuật đa hình tiểu vệ tinh được nhân bản ngẫu nhiên (Randomly Amplified Microsatellite Polymorphisms-RAMP)**

Các kỹ thuật chỉ thị dựa vào tiểu vệ tinh cho mức độ đa hình allele cao nhưng lại tốn nhiều công sức. Kỹ thuật RAPD không tốn kém nhưng mức độ đa hình thấp. Để dung hoàn yếu điểm của hai kỹ thuật này, kỹ thuật RAMP được xây dựng [201]. Kỹ thuật này sử dụng môi được đánh dấu chứa đầu 5' neo và đầu 3' lặp lại để nhân DNA genome với sự có mặt hoặc vắng mặt các môi RAPD. Sản phẩm nhận được phân tách trên gel polyacrylamide biến tính. Do môi



lặp lại được đánh dấu nên chỉ có sản phẩm của mỗi neo được xác định. Nhiệt độ bắt cặp của mỗi neo thường cao hơn mỗi RAPD 10-15°C. Vì vậy, ở nhiệt độ bắt mỗi cao chỉ có mỗi neo được kéo dài hiệu quả. Ở các chu kỳ PCR với nhiệt độ bắt mỗi thấp, cả hai mỗi neo và mỗi RAPD đều kéo dài. Do đó chương trình PCR được thiết lập sao cho có sự chuyển giữa nhiệt độ bắt mỗi cao xuống thấp trong quá trình phản ứng. Hầu hết các phân đoạn nhận được chỉ với các mỗi RAMP sẽ biến mất khi có các mỗi RAPD và các mẫu băng khác nhau nhận được bởi cùng một mỗi RAMP và các mỗi RAPD khác nhau cho thấy các mỗi RAPD cạnh tranh với mỗi RAMP trong chu kỳ với nhiệt độ bắt mỗi thấp. RAMP được sử dụng trong nghiên cứu đa dạng di truyền ở một số loài cây như lúa mạch [157, 201] và đào [33].

*Các kỹ thuật chỉ thị phân tử dựa vào gen nhảy (Transposable element-based molecular markers)*

Gen nhảy là các yếu tố di truyền có khả năng thay đổi vị trí của chúng trong hệ gen. Gen nhảy được phát hiện đầu tiên ở ngô [111]. Có hai loại gen nhảy. Loại I là gen nhảy ngược (retrotransposon), đây là yếu tố ngắn và dài nằm rải rác trong nhân, là yếu tố mã hóa mRNA và có mức thay đổi vị trí trung bình. Hiện tượng gen nhảy loại I tạo ra các bản sao mới trong khi bản gốc vẫn giữ nguyên ở vị trí ban đầu. Các bản sao được tạo ra theo hai giai đoạn: đầu tiên là phiên mã từ DNA thành RNA và tiếp theo là RNA tạo thành được phiên mã ngược thành DNA. Bản sao mới được gắn vào vị trí mới trong genome. Gen nhảy loại II (gen nhảy DNA) cấu tạo bởi gen nhảy DNA, loại này thay đổi vị trí bằng cơ chế “cắt và gắn” [56]. Chúng tự rời khỏi vị trí ban đầu gắn vào vị trí tiếp nhận. Dựa vào đặc điểm cấu trúc, gen nhảy còn phân chia thành dưới nhóm, trên họ, họ và dưới họ dựa trên sự định hướng của các khung đọc mở; vào sự có mặt, định hướng, độ dài và trình tự lặp lại bên trong và vào độ dài và trình tự nhân đôi ở vị trí đích tạo nên do sự thêm nucleotide [56].

Gen nhảy ngược (retrotransposons) hay gen nhảy loại I chia thành hai nhóm phụ khác nhau ở cấu trúc và chu kỳ chuyển vị đó là gen nhảy

có đầu lặp lại dài (Long Terminal Repeat (LTR) retrotransposon) và gen nhảy không có đầu lặp lại dài (non-LTR retrotransposon) bao gồm yếu tố LINE (Long INterspersed repetitive Element) và SINE (Short INterspersed repetitive Element). Hai nhóm phụ này được phân biệt bằng sự có mặt và vắng mặt đầu lặp lại dài (LTR) ở đầu. Tất cả các nhóm đều được bổ sung các dạng không phụ thuộc tương ứng thiếu một hoặc nhiều gen quan trọng cho sự chuyển vị như: các yếu tố có đầu lặp lại ngược nhỏ (Miniature Inverted repeat Terminal Elements-MITEs) cho nhóm II, SINEs cho non-LTR retrotransposons và gen nhảy ngược có đầu lặp lại nhỏ (Terminal-Repeat retrotransposons In Miniature-TRIMs) và phiên bản gen nhảy lùi lớn (large retrotransposon derivatives-LARDs) cho LTR retrotransposons.

Gen nhảy ngược (retrotransposons) cho phép tạo ra hệ chỉ thị phân tử tuyệt vời bởi chúng có chuỗi dài và bảo thủ, đa hình thêm nucleotide được tạo ra bởi hoạt động sao chép. Việc bổ sung nucleotide mới giúp cho việc tổ chức các hiện tượng bổ sung tạm thời trong dòng giống và như vậy có thể sử dụng để xác định phả hệ và phát sinh loài [63]. Phân tích phân tử dựa vào gen nhảy ngược được thực hiện với việc nhân bản bằng mỗi tương ứng với gen nhảy và mỗi phù hợp với vùng gen bên cạnh. Đa hình gen nhảy có một số kỹ thuật sau: Kỹ thuật đa hình chuỗi đặc thù được nhân bản (Sequence-Specific Amplified Polymorphism-S-SAP) là nhân bản DNA nằm giữa vị trí gắn của gen nhảy và vị trí cắt hạn chế cùng với chuỗi tiếp hợp [193]; kỹ thuật đa hình gen nhảy ngược giữa được nhân bản (Inter-Retrotransposon Amplified Polymorphism-IRAP) là nhân bản chuỗi DNA nằm giữa hai gen nhảy gắn nhau hoặc chuỗi lặp lại cuối trực tiếp dài (Long Terminal Direct Repeats-LTR) được nhân bản; kỹ thuật đa hình tiêu vệ tinh gen nhảy ngược được nhân bản (Retrotransposon-microsatellite Amplified Polymorphism-REMAP) được tiến hành bằng việc nhân bản phân đoạn nằm giữa vị trí gắn của gen nhảy và vị trí tiêu vệ tinh; kỹ thuật đa hình về vị trí gắn (insertion) dựa vào gen nhảy ngược (Retrotransposon-based Insertion Polymorphism-RBIP) được sử dụng để xác định các locus bị gen nhảy chiếm hoặc

locus rỗng; kỹ thuật biểu lộ gen nhảy (Transposable Display-TD).

Kỹ thuật S-SAP được sử dụng đầu tiên trong nghiên cứu vị trí của gen nhảy *BARE1* (barley retrotransposable element 1) trong lúa miến [93]. Về cơ bản, đây là dạng đơn giản của kỹ thuật AFLP. Việc nhân bản được thực hiện với mỗi đặc hiệu cho gen nhảy và mỗi đặc hiệu cho đoạn tiếp hợp *MseI* (*MseI*-adaptor). S-SAP được sử dụng trong nghiên cứu đa dạng di truyền và lập bản đồ [138, 144].

Trong kỹ thuật IRAP và REMAP đa hình được xác định bằng sự có mặt hoặc vắng mặt của gen nhảy ngược tại locus riêng biệt. Một phản ứng PCR có thể tạo được khoảng 30 băng. Các chỉ thị này có độ đa hình cao và có thể sử dụng trong đánh giá quan hệ trong loài. IRAP được sử dụng nghiên cứu đặc điểm hệ gen [126]. IRAP và REMAP được sử dụng để phát hiện sự giống nhau giữa các giống lúa [22].

Kỹ thuật RBIP được phát triển nhờ sử dụng gen nhảy *PDR1* của *Pisium sativum* [50]. Trong kỹ thuật này, cần phải có thông tin về trình tự đầu 5' và 3' của các vùng kề bên gen nhảy. Khi sử dụng mỗi đặc thù cho gen nhảy cùng với mỗi thiết kế để kéo dài tới vùng kề bên sẽ cho sản phẩm từ DNA khuôn có chứa nucleotide bổ sung. Mặt khác, các môi đặc thù cho cả hai vùng kề bên sẽ tạo sản phẩm nếu không có bổ sung nucleotide. Sự đa hình được xác định trên gel agarose hoặc lai với phân đoạn PCR tham khảo (tham chiếu). Đây là kỹ thuật khá tốn kém và phức tạp so với các kỹ thuật gen nhảy khác. RBIP được sử dụng trong nghiên cứu đa dạng di truyền và tiến hóa đậu đồng (*Pisum*) [77].

Kỹ thuật TD là dạng cải tiến của kỹ thuật AFLP, kỹ thuật này cho phép xác định đồng thời nhiều yếu tố phiên mã (TEs) từ các dòng có số lượng bản sao cao. Trong kỹ thuật này, sản phẩm PCR được tạo ra nhờ các môi neo ở vị trí cắt hạn chế của hai enzyme *BfaI* và *MseI* và yếu tố gen nhảy. Các gen nhảy được xác định bằng PCR gắn (ligation-mediated PCR) bắt đầu trong phạm vi gen nhảy và nhân bản từ phần chuỗi kề bên đến vị trí cắt hạn chế đặc thù. Sản phẩm PCR có thể phân tích trên gel polyacrylamide. TD được sử dụng đầu tiên để khám phá số bản sao của nhóm gen nhảy *dTph1* (TIRs) ở *Petunia*

và các sự kiện bổ sung nucleotide [189].

*Các kỹ thuật chỉ thị PCR các chuỗi đặc trưng*

*Kỹ thuật vùng nhân bản chuỗi được mô tả (Sequence Characterized Amplified Regions-SCAR)*

Để có thể sử dụng các chỉ thị xác định được bằng các môi ngẫu nhiên (như RAPD, AFLP v.v.) và khắc phục nhược điểm mắc cảm với điều kiện phản ứng của các phản ứng với môi ngẫu nhiên, kỹ thuật chỉ thị SCAR được phát triển và sử dụng. Kỹ thuật SCAR là kỹ thuật chỉ thị dựa vào PCR tạo ra các phân đoạn DNA tại các locus xác định bằng sử dụng các môi nucleotide đặc thù [137]. SCAR được tiến hành bằng cách tách dòng sản phẩm PCR với các môi ngẫu nhiên sau đó đọc trình tự hai đầu đoạn tách dòng. Dựa vào trình tự hai đầu này để thiết kế cặp môi với độ dài 15 đến 30 bp để nhân băng đơn với kích thước tương tự như phân đoạn được nhân dòng. Sự đa hình được xác định bằng sự có mặt hay vắng mặt của băng được nhân bản hoặc sự suất hiện đa hình về chiều dài đồng thời chuyển chỉ thị trội thành chỉ thị SCAR đồng trội. Kỹ thuật SCAR được dùng để phân tích thư viện genome [31], xác định các locus đặc thù [137], chọn giống nhờ chỉ thị phân tử [95].

*Kỹ thuật vị trí chuỗi đánh dấu (Sequence-tagged Sites-STS)*

Chuỗi vị trí đánh dấu (STS) là vùng ngắn trong genome (có độ dài 200-300 cặp bazơ) mà trình tự của nó không thể có ở bất cứ nơi nào khác trong genome. Chuỗi trình tự duy nhất này có thể nhân bản được bằng PCR. Trình tự DNA của STS có thể có các yếu tố lặp lại, và các trình tự xuất hiện bất cứ ở đâu trong genome, nhưng trình tự ở hai đầu của STS là duy nhất. Chúng ta có thể tổng hợp các môi duy nhất bổ sung với trình tự ở hai đầu STS và nhân đoạn STS bằng PCR. Với nghĩa rộng, STS bao gồm các chỉ thị như microsatellites (SSRs, STMS or SSRPs), SCARs, CAPs và ISSRs.

Kỹ thuật STS là sử dụng PCR để nhân chọn lọc vị trí đánh dấu (STS) bằng cặp môi đặc hiệu gắn với hai đầu của chuỗi STS với điều kiện của phản ứng sao cho chỉ nhân được chuỗi STS đích mà không phải chuỗi nào khác trong genome.

Sản phẩm PCR được phân tách trên gel agarose hoặc polyacrylamide. Ưu điểm của kỹ thuật STS là tạo ra các chỉ thị đồng trội, có thể phân biệt được dị hợp tử và có tính lặp lại cao vì các môi sử dụng dài. Nhược điểm của kỹ thuật này là cần biết về trình tự vùng STS và cần đầu tư cho việc phát triển mỗi đặc hiệu.

STS được sử dụng cho nghiên cứu đa dạng di truyền [139], nhận biết gen [89], so sánh bản đồ [25], xác định gen liên kết [125], xác định độ thuần hạt lai [206] và đánh giá tính kháng bệnh [13].

*Các kỹ thuật chỉ thị dựa trên phản ứng PCR khác*

Kỹ thuật các chuỗi đa hình nhân bản được cắt (Cleaved Amplified Polymorphic Sequences-CAPS)

Kỹ thuật chỉ thị CAPS là giải pháp sử dụng các chuỗi DNA của chỉ thị RFLP đã được lập bản đồ để phát triển chỉ thị PCR nhằm tránh bước lai Southern phức tạp. Vì thế, kỹ thuật CAPS còn được gọi là kỹ thuật PCR-RFLP [93]. Kỹ thuật CAPS được tiến hành bằng cắt hạn chế sản phẩm PCR của các locus đặc thù với một hoặc nhiều enzyme cắt hạn chế và phân tách các đoạn DNA được cắt trên gel agarose hoặc polyacrylamide. Các cặp môi được thiết kế dựa trên thông tin về trình tự của DNA hoặc cDNA trên ngân hàng gen hoặc trình tự các băng RAPD được tách dòng. Chỉ thị CAPS là chỉ thị đồng trội và là các locus đặc thù được dùng để phân biệt đồng hợp tử và dị hợp tử ở các loài thực vật [93]. Kỹ thuật CAPS có hạn chế vì sự đột biến có thể làm mất hoặc tạo thêm các vị trí nhận biết của các enzyme cắt hạn chế. Để khắc phục hạn chế này, Michae & Amasino (1998) [118] đưa ra kỹ thuật biến thể khác gọi là dCAPS. Trong phân tích dCAPS, vị trí nhận biết của enzyme cắt hạn chế có SNP được đưa vào sản phẩm PCR bằng môi chứa một hoặc nhiều điểm bất hợp (mis-match) với khuôn DNA. Sản phẩm PCR cải tiến này sau đó được cắt bằng enzyme cắt hạn chế và sự có mặt hay vắng mặt của SNP được xác định bằng mẫu cắt hạn chế tạo ra. CAPS và dCAPS là phương pháp xác định SNP cắt sản phẩm PCR đặc biệt bằng các enzyme cắt hạn chế. Đây là kỹ thuật rẻ thực hiện, ít tốn kém so với bất kỳ phương pháp

nào. Kỹ thuật này rất hữu ích cho việc theo dõi các đột biến đã biết trong quần thể phân ly [128] và để phát triển chỉ thị, đặc biệt là chỉ thị SNP liên kết với gen/QTL phục vụ cho chọn giống [40, 100].

*Kỹ thuật đa hình các chuỗi liên quan được nhân bản (Sequence-Related Amplified Polymorphism-SRAP)*

Kỹ thuật SRAP [102] là kỹ thuật nhân bản các khung đọc mở (ORFs). Trong kỹ thuật này dùng hai môi ngẫu nhiên có độ dài 17 đến 21 nucleotide và giàu AT hoặc GC để nhân phân đoạn bên trong gen cho đa hình đã được xác định. Các môi cấu tạo bởi các thành phần: các chuỗi lõi (core primers) với 13-14 nucleotide, trong đó 10-11 nucleotide đầu bắt đầu từ đầu 5' là trình tự không đặc thù (hay còn gọi là chuỗi rời) tiếp theo là các trình tự CCGG cho môi xuôi và AATT cho môi ngược. Các môi lõi được gắn với ba nucleotide chọn lọc ở đầu 3'. Chuỗi rời của mỗi ngược và xuôi phải khác nhau và có thể dài 10-11 nucleotide. Năm chu kỳ đầu với nhiệt độ kéo dài mỗi là 35°C, 35 chu kỳ tiếp theo nhiệt độ kéo dài mỗi là 50°C. Các phân đoạn DNA nhân bản được phân tích bằng gel acrylamide biến tính. Kỹ thuật SRAP kết hợp sự đơn giản, độ tin cậy và dễ dàng đọc trình tự các băng chọn lọc. SRAP hướng tới chuỗi mã hóa trong hệ gen tạo ra các chỉ thị đồng trội. Đa hình các chuỗi nhân bản tạo nên do hai sự kiện là sự thay đổi độ dài phân đoạn DNA do bổ sung hoặc mất nucleotide tạo ra chỉ thị đồng trội và sự thay đổi nucleotide tạo ra chỉ thị trội. SRAP được dùng để lập bản đồ [34, 106, 166] và nghiên cứu đa dạng di truyền [59, 163].

*Kỹ thuật đa hình cấu tạo sợi đơn (Single Strand Conformation Polymorphism-SSCP)*

Đa hình cấu tạo sợi đơn là kỹ thuật cải tiến phân tích sự thay đổi vị trí của các chuỗi DNA đơn trên điện di gel polyacrylamide trung tính để xác định đa hình tạo nên bởi sự cuộn khác nhau của sợi đơn DNA do các thay đổi khó thấy trong chuỗi nucleotide [136]. Phân tích SSCP là công cụ mạnh cho việc đánh giá sản phẩm PCR phức tạp khi hai sợi DNA từ cùng sản phẩm PCR thường chạy tách rời trên SSCP gel, vì thế, có hai khả năng ghi nhận đa hình và giải quyết vấn đề đa hình bên trong chuỗi của sản phẩm

PCR từ vùng giống hệt nhau trong hệ gen của các cá thể nghiên cứu. Nguyên lý của phương pháp phân tích SSCP là dựa trên thực tế: sợi đơn DNA có cấu tạo nhất định. Thay đổi cấu tạo bằng sự thay đổi một nucleotide sẽ làm cho sợi đơn DNA chuyển động khác đi trong điều kiện điện di trên gel không biến tính. Vì thế, mẫu nguyên thủy và mẫu đột biến sẽ có mẫu điện di đồ khác nhau. Phân tích SSCP được tiến hành theo các bước: i. nhân bản PCR chuỗi DNA quan tâm, ii. biến tính sản phẩm PCR, iii. làm lạnh chuỗi DNA biến tính để tạo sợi đơn, iv. xác định sự di chuyển khác nhau giữa các sợi đơn trong điều kiện điện di trên gel không biến tính. Có thể quan sát sự khác nhau bằng đánh dấu phóng xạ, nhuộm bạc, bằng các mồi PCR được đánh dấu bằng màu huỳnh quang và điện di mao quản. PCR-SSCP là kỹ thuật nhanh, đơn giản, nhạy dùng cho xác định các đột biến bổ sung, thêm vào và bớt đi nucleotide trong phân đoạn DNA được nhân bản. Đây là công cụ quan trọng cho phân tích gen đặc biệt là xác định đột biến điểm. Kỹ thuật này giống kỹ thuật RFLP vì nó có thể giải đoán những thay đổi allele của các tính trạng di truyền. Khác với RFLP là SSCP có thể xác định được đa hình DNA và đột biến ở nhiều chỗ trong phân đoạn DNA. PCR-SSCP huỳnh quang là dạng kỹ thuật thích ứng của SSCP trong đó sự nhân bản chuỗi DNA đích được tiến hành với mồi đánh dấu huỳnh quang. Sự hạn chế của kỹ thuật SSCP là việc xây dựng chỉ thị SSCP mất nhiều công sức, đắt và không tự động hóa được. Kỹ thuật SSCP được xây dựng để xác định đột biến [136] sau đó được sử dụng để xác định biến đổi của chuỗi DNA [162] và nhân dòng gen [204].

*Kỹ thuật đa hình nucleotide đơn (Single Nucleotide Polymorphism-SNP)*

Những thay đổi một nucleotide trong trình tự genome của các cá thể trong quần thể được gọi là đa hình nucleotide đơn (SNP). Các vị trí thể hiện SNP trong genome là nơi mà ở đó chuỗi DNA được phân biệt bởi một bazơ duy nhất khi hai hoặc nhiều cá thể được so sánh. Sự khác nhau về nucleotide này có thể dẫn đến thay đổi tính trạng đặc biệt hay kiểu hình, hoặc có thể là sự thay đổi trung tính có thể được sử dụng để đánh giá sự đa dạng trong tiến hóa. SNP là dạng thay đổi trình tự trong genome phổ biến

nhất cho tới nay. Ở ngô cứ 60 đến 100 bp có một SNP [35], ở người 90% thay đổi trình tự là thay đổi nucleotide đơn và cứ 1000 bp có một SNP [155].

Ở thực vật, SNP đang được thay thế cho SSR như chỉ thị cho các ứng dụng trong di truyền và chọn giống. SNP có số lượng lớn, ổn định, hiệu quả, cho phép tự động hóa, và ngày càng kinh tế hơn [44, 46]. Hơn nữa, SNP xảy ra cả ở các vùng mã và không mã của DNA nhân và lục lạp. Nguồn SNP hiện đang được phát triển và thông tin rộng rãi cho nhiều ứng dụng ở lúa. Các nguồn này bao gồm cơ sở dữ liệu SNP, phương tiện để xác định các SNP mang nhiều thông tin cho các mục đích ứng dụng và bộ SNP cho đánh giá được thiết kế theo đặt hàng phục vụ mục đích chọn lọc nhờ chỉ thị phân tử.

Mặc dù SNP có các ưu việt hơn các kỹ thuật khác, nó vẫn có một số nhược điểm như: sự hạn chế khám phá các SNP trong các cơ thể không phải hình mẫu do sự tồn kém và khó khăn trong các công nghệ đang được sử dụng để phát hiện SNP.

Có rất nhiều phương pháp để phát hiện SNP bao gồm các phương pháp lai (lai allele đặc biệt, SNP microarray), các phương pháp sử dụng các enzyme (RFLP, PCR, phương pháp kéo dài mồi, sử dụng 5'-nuclease v.v.), các phương pháp dựa trên tính chất vật lý của DNA đối với các sản phẩm PCR (SSCP, điện di gradient nhiệt độ, sắc ký lỏng cao áp biến tính v.v.) và phương pháp giải trình tự. Độ giá có thể tìm hiểu sâu hơn về các phương pháp phát hiện SNP trong công bố của Kwork & Chen (2003) [94] và tổng quan của Jehan & Lakhanpaul (2006) [76].

Chiến lược trực tiếp điển hình cho khám phá các SNP là giải trình tự các sản phẩm nhân bản các locus đặc thù (locus specific amplification-LSA) từ nhiều cá thể hoặc xác định trình tự các chuỗi thể hiện được đánh dấu (giải trình tự các EST) [165, 188]. Các chiến lược trực tiếp khác có thể kể đến như: giải trình tự toàn bộ hệ gen hoặc giải trình tự các vùng đại diện. Nếu có các dữ liệu về các chuỗi cần so sánh trên mạng thông tin đại chúng hoặc các cơ sở dữ liệu khác thì có thể tiến hành các so sánh để phát hiện ra SNP [61]. Phân tích trực tiếp sự

khác nhau trong trình tự giữa nhiều cá thể với số lượng lớn các locus có thể đạt được bởi công nghệ giải trình tự thế hệ thứ hai (next-generation sequencing). Giải trình tự lại hay tái giải trình tự (re-sequencing) được sử dụng để xác định sự thay đổi ở các cá thể có thể cho các chỉ thị di truyền phân tử và thấy được vai trò của gen. Quá trình tái giải trình tự toàn bộ genome (whole-genome re-sequencing) sử dụng công nghệ đọc ngắn (short-read technology) bao gồm sự so sánh những bộ hàng triệu chuỗi đọc lần lượt từng nucleotide với chuỗi genome so sánh. Bằng kỹ thuật này có thể xác định được sự biến đổi trong chuỗi nucleotide của mẫu nghiên cứu và đối chứng.

SNP được sử dụng trong lập bản đồ di truyền [88, 146, 203], trong nghiên cứu đa dạng di truyền [53, 83], trong nhận biết giống [74] và trong chọn giống [62].

*Kỹ thuật phân tích vùng đệm trong trong được sao mã (Internal Transcribed Spacer-ITS)*

Các gen RNA ribosome (hay ribosome DNA-rDNA) là các phần của các đơn vị lặp lại được sắp xếp theo thứ tự. Chúng được tìm thấy ở các vị trí nhiễm sắc thể được gọi là các vùng tổ chức nhân (nucleolar organizing regions-NORs). DNA ribosome nhân có hai vùng sao mã: ITS-1 nằm giữa tiểu phần nhỏ (16S-18S) và 5,8S của gen và ITS-2 nằm giữa 5,8S và tiểu phần lớn (23S-28S) của gen. Hai vùng này và tiểu phần 5,8S được gọi chung là vùng sao mã bên trong (ITS).

Các vùng ITS có độ dài 600 đến 700 bp là các vùng tiến hóa nhanh nên có thể thay đổi về trình tự cũng như độ dài. Các vùng bên cạnh ITS lại rất bảo thủ nên được sử dụng để thiết kế các mối chung cho nhân bản vùng ITS.

Số bản sao các đoạn lặp lại của rDNA lên tới 30 000 trong một tế bào. Điều này làm cho ITS trở thành đối tượng lý thú cho nghiên cứu tiến hóa, phát sinh loài [11] và đa dạng di truyền [140]. ITS là kỹ thuật quan trọng cho phân loại phân tử các nhóm phân loại (taxon) có mối liên kết gần, bởi vì ITS có tính bảo thủ cao trong loài nhưng lại thay đổi ở các loài khác nhau [26]. Các nghiên cứu về phát sinh loài dựa vào nrDNA hay các chuỗi ITS cho phép hiểu biết sâu về tiến hóa và sự lai tạo ở các loài thực vật

khác nhau.

Nghiên cứu đa dạng di truyền sử dụng ITS có thể tiến hành bằng việc giải trình tự trực tiếp vùng quan tâm từ các cá thể khác nhau sau đó tiến hành xây dựng cây phân loại dựa vào số liệu so sánh các trình tự. Cũng có thể xác định sự thay đổi của các chuỗi bằng cắt hạn chế vùng ITS bằng các enzyme cắt hạn chế và phân tách các phân đoạn bằng gel agarose hoặc polyacrylamide. Những thay đổi về các chuỗi tạo ra bằng kỹ thuật cắt hạn chế vùng ITS hay còn gọi là đa hình độ dài đoạn cắt giới hạn (PCR-RFLP) có thể dùng trong phân loại.

Do sự bảo thủ trong cấu tạo bậc hai, ITS được xem như công cụ sắc bén cho việc so sánh tiến hóa của cơ thể nhân thực [184]. Sở dĩ vậy là do vùng ITS bảo thủ cao trong loài nhưng lại thay đổi ở các loài khác nhau. Đối với nhiều cơ thể, ITS2 trong rRNA được tổ chức xung quanh trung tâm bảo thủ chung của cấu trúc bậc hai từ đó bốn vòng xoắn được tạo ra. Cơ sở dữ liệu về cấu trúc của ITS2 chứa hơn 288.000 cấu trúc tiên đoán cho các chuỗi ITS2 đã biết hiện nay. Các trình tự này của ITS2 mở ra một khả năng mới là kết hợp các thông tin về cấu trúc trong nghiên cứu tiến hóa.

ITS được sử dụng trong nghiên cứu quan hệ di truyền và phân loại [151, [211], trong xác định con lai [148, 192], trong nghiên cứu nguồn gốc, phát sinh loài [90] và trong nghiên cứu mã vạch thực vật [207].

*Các kỹ thuật chỉ thị DNA lục lạp*

DNA tế bào chất gồm DNA lục lạp và DNA ty thể. Các chỉ thị DNA lục lạp cũng như ty thể trở nên rất hiệu quả trong nghiên cứu đa dạng di truyền và phân loại. Phân tích DNA lục lạp cung cấp những thông tin về đa dạng di truyền thực vật tương đồng với các DNA nhân. Hệ gen lục lạp có tính bảo thủ cao và mức độ biến thấp hơn so với hệ gen nhân. Vì vậy, chỉ thị lục lạp được sử dụng rộng rãi trong nghiên cứu đa dạng di truyền. Các kỹ thuật chỉ thị DNA lục lạp được sử dụng bao gồm: phân tích vị trí cắt hạn chế của DNA lục lạp (cpDNA), phân tích các RNA vận chuyển (tRNA) lục lạp, phân tích SSR lục lạp (cpSSR) và phân tích trình tự cpDNA.

*Kỹ thuật phân tích vị trí cắt hạn chế DNA lục lạp (RFLP cpDNA)*

Phân tích cắt hạn chế DNA lục lạp (cpDNA) được sử dụng rộng rãi trong nghiên cứu các loài và các biến đổi trong loài. Phân tích sự thay đổi độ dài các đoạn cắt hạn chế cpDNA là phương pháp đầu tiên trong phân tích cpDNA cho mục đích nghiên cứu phát sinh loài. Phương pháp phân tích cắt hạn chế cpDNA bao gồm phương pháp đơn giản như so sánh toàn bộ các đoạn cắt hạn chế genome lục lạp trên gel agarose đến lập bản đồ so sánh trực tiếp các vị trí cắt hạn chế bằng lai Southern. Phân tích điểm cắt hạn chế cpDNA có nhiều ưu việt trong nghiên cứu phát sinh loài: i- kỹ thuật đơn giản so với nhiều kỹ thuật phân tử khác; ii- hệ gen lục lạp đủ lớn để có thể có được nhiều vị trí cắt cho một enzyme; iii- các vị trí cắt hạn chế thể hiện gần như mẫu ngẫu nhiên của cpDNA; iv- các thay đổi mang thông tin phá hệ xảy ra ở các vị trí nằm trên vùng không mã; v- phân tích điểm cắt hạn chế không bị ảnh hưởng bởi sự nhiễm như đối với giải mã sản phẩm PCR. Tuy nhiên, do tính bảo thủ của cpDNA nên các loài gần nhau thường khó có đa hình. Phân tích vị trí cắt giới hạn cpDNA cũng được thực hiện trên các chuỗi gen (*matK*, *atpB*, *ndhF*, *rbcL*), các vùng không mã (*ndhF* và *trnL* intron) và các vùng đệm giữa các gen lục lạp (*trnH-trnK*, *psbC-trnS*, *trnD-trnS*, *trnT-trnF*, *trnL-trnF* v.v.) bằng việc cắt sản phẩm PCR nhân bản các gen và các vùng này bằng các enzyme cắt hạn chế. Phương pháp này còn được biết đến là phương pháp PCR-RFLP cpDNA. Phân tích vị trí cắt hạn chế là phương pháp so sánh gián tiếp sự thay đổi trong genome. Để so sánh trực tiếp cần tiến hành đọc trình tự gen sau đó so sánh.

Phân tích cắt hạn chế cpDNA được sử dụng trong nghiên cứu tiến hóa, phân loại, địa sinh học [133], nghiên cứu đa dạng di truyền và biến đổi cpDNA [141, 174, 176, 177, 180].

#### Kỹ thuật phân tích các RNA vận chuyển lục lạp

Các RNA vận chuyển (tRNAs) lục lạp là các đại phân tử cổ đã tiến hóa dưới các áp lực môi trường như các yếu tố thích ứng trong sự chuyển đổi ở tất cả các dạng sống, nhưng cũng dẫn tới các thay đổi cấu trúc và chức năng. Nói cách khác tRNAs rất đa dạng về vai trò (tổng hợp thành tế bào, tổng hợp porphyrin cho chlorophyll, biến đổi đầu N của protein, khởi

động phiên mã ngược, tái mô hình hóa lipid trong sự bổ sung vào tổng hợp protein phụ thuộc ribosome) và cấu trúc. Một đặc điểm của các vùng tRNA là rất bảo thủ và vì thế cho phép những thay đổi cấu trúc dưới dạng thay thế, thêm bớt nucleotide (indel-insertion and deletion) và chuỗi lặp lại. Những thay đổi này cung cấp những thông tin quan trọng về sự tiến hóa.

Trong phân tích các RNA vận chuyển, các công trình nghiên cứu thường sử dụng các vùng đệm giữa các gen, đặc biệt là giữa hai gen *trnL-trnF*. Vùng đệm giữa hai gen *trnL-trnF* có độ dài nhỏ hơn 500 bp. Từ vùng bảo thủ, hai môi đã được thiết kế để nhân vùng đệm này ở các loài [167]. Với các môi chung này và mức độ đa hình cao trong vùng đệm *trnL-trnF* làm cho vùng này trở thành chỉ thị tốt cho nghiên cứu phân tích ở nhiều loài thực vật [32]. Các trình tự khác nhau giữa các loài trong vùng đệm này có thể xác định bằng điện di trên gel polyacrylamide, phản ứng PCR-SSCP hoặc thậm chí điện di trên gel agarose. Ngoài vùng đệm giữa gen *trnL-trnF*, một số vùng đệm khác như các vùng đệm giữa các gen *trnH-trnK*, *psbC-trnS*, *trnD-trnS*, *trnT-trnF* cũng được sử dụng [129]. Các vùng không mã (intron) của gen *trnL* (UAA) cũng được sử dụng nhiều cho phân tích tiến hóa thực vật và cho phát triển chỉ thị để xác định thực vật cho gen là mẹ ở cơ thể đa bội cùng với khả năng khám phá quan hệ tiến hóa của các loài liên quan [205].

Kỹ thuật Simple sequence repeats lục lạp (cpSSR)

SSR lục lạp (cpSSR) hay tiểu vệ tinh lục lạp (chloroplast microsatellite) được tìm thấy trong tất cả các hệ gen lục lạp đã được giải trình tự hoàn toàn và hàng trăm trình tự lục lạp chưa được giải trình tự hết. Việc nhân bản bằng PCR các cpSSR này bằng các môi đặc thù với các vùng nằm bên đầu 5' và đầu 3' đã tạo ra các sản phẩm có độ dài khác nhau tương ứng với vùng cpSSR. SSR lục lạp (cpSSR) khác với SSR nhân (nSSR) ở chỗ cpSSR cấu tạo bởi kiểu nucleotide đơn lặp lại từ 8 đến 15 lần. cpSSR tiến hóa nhanh hơn các vùng gen khác trong hệ gen lục lạp [75] và được xác định ở nhiều loài cây. Cho đến nay, các nghiên cứu cho thấy

cpSSR cho mức độ đa hình cao hơn phân tích cắt hạn chế cpDNA (RFLP-cpDNA) [142].

Hệ gen lục lạp được di truyền với nhiều bản sao trong phân chia mitosis và meiosis, vì vậy, nó trở thành đối tượng nghiên cứu phân hóa do trôi gen trong cấu trúc di truyền và trôi gen ngoài tự nhiên [65] và để xác định sự suy giảm đa dạng di truyền [116]. Di truyền theo mẹ, đơn bội, và bản chất không tái tổ hợp của hệ gen lục lạp làm nó trở thành phương tiện hữu hiệu trong nghiên cứu tiến hóa. cpSSR đã được sử dụng trong nghiên cứu nhiều loài cây như lúa, lúa mạch, thông, gỗ đỏ v.v. [45, 142, 143, 180].

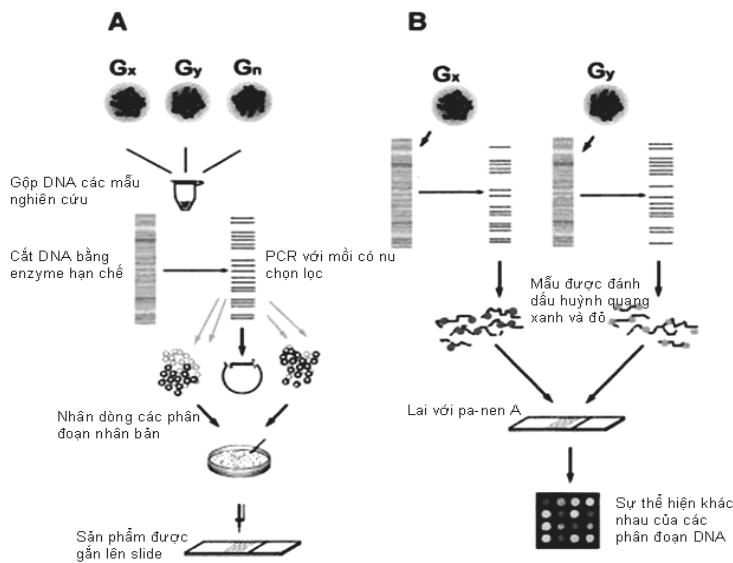
#### Kỹ thuật phân tích trình tự cpDNA

Trình tự DNA cung cấp công cụ cho phân tích so sánh trực tiếp. Tiêu chí chọn trình tự DNA cho phân tích bao gồm: i. trình tự phải đủ dài để có đủ thông tin về tiến hóa di truyền các vị trí nucleotide, ii. mức độ đa dạng phù hợp với vấn đề tiến hóa di truyền được đặt ra, theo Ritland và Clegg [151] mức độ đa dạng từ 5 đến 15% là phù hợp, mức độ này cung cấp đủ số

lượng đặc điểm cần thiết. Phân tích trình tự cpDNA có ưu điểm là cần rất ít cpDNA khuôn. Để phân tích trình tự cpDNA, cũng giống như phân tích trình tự DNA nói chung, trước hết là nhân các gen, vùng gen hay vùng đệm các gen lục lạp bằng các môi đặc hiệu, sau đó có thể nhân dòng trước khi đọc trình tự hoặc đọc trình tự trực tiếp sản phẩm PCR. Cuối cùng là so sánh trình tự bằng các phần mềm chuyên dụng.

Nhiều nghiên cứu đã sử dụng trình tự gen *rbcL* cho nghiên cứu phân loại và tiến hóa loài [132] một số khác sử dụng trình tự gen *ndhF* [55] và gen *matK* [91]. Trình tự các chuỗi vùng đệm giữa các gen hay trình tự các chuỗi không mã của các gen hoặc giữa hai gen (*trnL* (UAA), *rbcL-atpB* cũng được sử dụng [130, 208]. Trình tự các vùng gen *rbcL*, *matK*, *atpF-atpH* IGS, *psbK-psbI* IGS và *trnH-psbA* IGS còn được dùng cho mã vạch (barcode) ở các loài lan [87].

#### Công nghệ thể hiện đa dạng (Diversity Arrays Technology-DArT)



Hình 4. Sơ đồ các bước tiến hành DArT

A. Pa-nen đa dạng. Genomic DNA của các mẫu nghiên cứu được gộp với nhau. Sau đó DNA được cắt bằng các enzyme cắt hạn chế được chọn và gắn với các tiếp hợp. Mức độ phức tạp của genome được giảm thiểu bằng PCR với các môi có nucleotide chọn lọc. Các phân đoạn đại diện được nhân bản bằng các môi đặc hiệu cho vector và nhân dòng, sản phẩm PCR được tinh sạch và gắn lên slide.

B. Sự khác nhau giữa hai mẫu thể hiện bằng DArT. Hai mẫu genome được chuẩn bị theo phương pháp mô tả ở (A). Một mẫu được đánh dấu huỳnh quang màu xanh, một mẫu được đánh dấu màu đỏ, trộn với nhau và lai với pa-nen đa dạng. Tỷ lệ mức độ tín hiệu xanh/đỏ được đo ở mỗi lần và sự khác nhau về tỷ lệ tín hiệu cho thấy sự thể hiện khác nhau của các phân đoạn DNA [73].

Công nghệ thể hiện đa dạng (DArT) là công nghệ phân tích kiểu gen có tính phù hợp rộng và hiệu quả về kinh tế. Công nghệ này được phát

triển để khắc phục một số nhược điểm của các công nghệ chỉ thị phân tử khác như RFLP, AFLP và SSR [4]. Đây là công nghệ khắc phục

nhược điểm tốn kém về thời gian của các kỹ thuật dựa trên việc lai DNA và có thể phân tích hàng ngàn locus trong một lần thí nghiệm. DArT đặc biệt phù hợp cho phân tích kiểu gen các loài đa bội với hệ gen lớn. Công nghệ này có thể lấy dấu (fingerprint) toàn bộ hệ gen bằng việc ghi nhận sự có mặt hoặc vắng mặt các phân đoạn DNA trong hệ gen đại diện được hình thành từ các mẫu DNA genome. Công nghệ DArT gồm nhiều bước như: giảm thiểu sự phức tạp của mẫu DNA (bằng gộp DNA của các mẫu, cắt DNA bằng enzyme giới hạn và PCR với các môi có nucleotide chọn lọc), tạo thư viện DNA, đưa các thư viện lên bản kính, lai với các DNA được đánh dấu huỳnh quang trên bản kính, quét bản kính để xác định tín hiệu lai, ghi chép số liệu và phân tích (hình 4).

DArT giảm thiểu sự phức tạp của mẫu DNA để nhận đại diện của mẫu. Sự giảm thiểu này dựa trên việc kết hợp cắt hạn chế và gắn chuỗi tiếp hợp và kèm theo sự nhân bản bằng PCR. Chỉ thị DArT cho loài mới được khám phá bằng việc sàng lọc thư viện hàng ngàn phân đoạn DNA từ genome đại diện được chuẩn bị từ việc gộp các mẫu DNA chứa đựng sự đa dạng của loài. Hệ thống phân tích chip vi điểm (microarray platform) làm cho quá trình khám phá hiệu quả vì tất cả các chỉ thị trên một phân tích DArT cụ thể được ghi nhận đồng thời. Đối với mỗi phương pháp giảm thiểu phức tạp có một bộ chỉ thị DArT riêng phù hợp cho phân tích DArT riêng rẽ. Do đó, số lượng các chỉ thị cho mỗi loài nhất định chỉ phụ thuộc vào mức độ biến đổi trong loài và số phương pháp giảm thiểu phức tạp được sàng lọc.

#### ***Công nghệ giải trình tự thế hệ thứ hai (Next-Generation of Sequencing Technology)***

Sự phát triển các công nghệ giải trình tự số lượng lớn (high throughput sequencing) đã làm cho việc giải trình tự DNA trở nên đặc biệt quan trọng cho các nghiên cứu về đa dạng di truyền, tiến hóa, bảo tồn sinh học và chọn giống. Các công nghệ này có tiềm năng loại bỏ một trong các trở ngại đối với thực hiện tiếp cận hệ gen trong các cơ thể không phải là mô hình bao gồm các cơ thể thiếu thông tin về trình tự genome. Các công nghệ này tránh được sự tốn kém, phức tạp và sai lệch ở phương pháp giải trình tự dựa

vào nhân dòng bằng nhân bản trực tiếp từ DNA khuôn [17, 113].

Công nghệ giải trình tự thế hệ thứ hai (hay còn được coi là công nghệ giải trình tự số lượng lớn) gồm ba nhóm chính là giải trình tự bằng phương pháp tổng hợp, giải trình tự bằng phương pháp gắn và giải trình tự phân tử đơn.

*Giải trình tự bằng phương pháp tổng hợp* giống như kỹ thuật giải trình tự của Sanger tức là xác định thành phần nucleotide bằng phát hiện tín hiệu quang hóa trong quá trình gắn các nucleotide vào sợi DNA bổ trợ bằng enzyme DNA polymerase. Trong kỹ thuật Sanger, sự kết thúc chuỗi dideoxynucleotide được dùng để xác định trình tự, còn trong giải trình tự bằng phương pháp tổng hợp thì DNA được cắt thành nhiều phân đoạn rồi gắn với chuỗi tiếp hợp sau đó nhân dòng để tăng tín hiệu huỳnh quang hoặc tín hiệu hóa học. Có ba hệ thống giải trình tự bằng phương pháp tổng hợp. Ba hệ thống này khác nhau về độ dài đọc, về phương pháp nhân dòng và cố định, bao gồm: hệ thống 454 của Roche, Solexa của Illumina và SOLiD của Applied Biosystem.

Hệ thống Roche 454 giải trình tự bằng phương pháp dựa trên nguyên tắc tổng hợp cao nhiệt (pyrosequencing), sợi đơn DNA khuôn được gắn vào hạt siêu nhỏ (microbead) và nhân bản bằng PCR nhũ tương (emulsion PCR) [112]. PCR nhũ tương là phản ứng PCR sử dụng hỗn hợp phản ứng PCR thông thường và hỗn hợp phản ứng PCR dạng nhũ tương nước trong dầu [158]. Hệ thống Roche 454 đầu tiên chỉ đọc được đoạn 100 bp, hiện nay do cải tiến có thể đọc đoạn đến 800 bp.

Hệ thống Illumina của Solexa giải trình tự dựa vào việc đơn giản hóa phương pháp xây dựng thư viện và đảo đầu bằng phương pháp huỳnh quang dẫn đến việc tạo ra các đoạn đọc có độ dài 35 bp [17]. Hệ thống này sử dụng việc nhân bản cầu pha cứng (solid-phase bridge amplification), trong đó các đoạn tiếp hợp đầu 5' và 3' được gắn với mỗi đầu của khuôn DNA. Một đầu của khuôn sau đó được dính với cơ chất. Các đoạn tiếp hợp được lai với các môi xuôi hoặc ngược cố định tạo thành cầu nối, tạo thuận lợi cho việc nhân bản để tạo ra các sản phẩm nhân bản được dính với cơ chất và tạo



thành các nhóm khuôn giống nhau do đó làm tăng cường sự phát hiện bằng quang hóa. Với hệ thống HiSeq 2000 có thể tạo khoảng 6 tỷ đoạn đọc cho 540 đến 600 Gb trong 11 ngày (<http://www.illumina.com>).

Hệ thống Ion Torrent (<http://www.iontorrent.com>) là hệ thống giải trình tự thế hệ thứ hai duy nhất xác định trình tự không dựa vào các chất huỳnh quang mà dựa vào việc đo sự thay đổi pH do việc giải phóng H<sup>+</sup> khi gắn nucleotide bằng công nghệ bán dẫn (semiconductor) [154]. Bằng việc bổ sung liên tục các nucleotide, máy có thể nhận biết nucleotide nào được gắn vào sợi đang kéo dài.

*Giải trình tự bằng phương pháp gắn* là giải trình tự bằng tổng hợp có sử dụng DNA polymerase làm phương tiện kéo dài trong quá trình xác định trình tự DNA. Giải trình tự bằng phương pháp gắn khai thác sự mẫn cảm của DNA ligase với sự bất cập sai (mismatch) để xác định trình tự DNA. Phương pháp này sử dụng các oligonucleotide dò có kích thước khác nhau và được đánh dấu bằng chất huỳnh quang ở nucleotide cận xác định, các phân đoạn DNA khuôn được nối với các chuỗi neo ngắn đã biết để tạo điều kiện cho sự lai với các đoạn dò. DNA ligase được bổ sung vào phản ứng để nối các đoạn dò được đánh dấu huỳnh quang với mỗi và khuôn. Hình ảnh huỳnh quang được thiết lập để xác định đoạn dò nào được gắn. Quá trình này được lặp lại với việc sử dụng các bộ dò khác nhau cho các DNA khuôn nghiên cứu để xác định trình tự nucleotide. Hệ thống giải trình tự hỗ trợ gắn và phát hiện oligonucleotide (Supported Oligonucleotide Ligation and Detection-SOLiD) của Life Technologies/Applied Biosystems (<http://www.appliedbiosystems.com>) giải trình tự bằng phương pháp gắn có thể tạo ra chuỗi DNA 0,1 đến 4 Gb trong một đến bảy ngày với giá thành trong khoảng 3400 đến 8500 Đô-la Mỹ.

*Giải trình tự phân tử đơn* còn gọi là *giải trình tự thế hệ thứ ba* (third-generation sequencing). Phương pháp này tạo ra các tín hiệu nhận biết sự gắn nucleotide bằng quang hóa trong quá trình giải trình tự từ phân tử nucleic acid đơn. Vì thế có thể loại bỏ việc nhân bản khuôn. Điều này làm cho giải trình tự phân

tử đơn có nhiều lợi thế so với giải trình tự thế hệ thứ hai. Đặc biệt là việc đơn giản hóa sự chuẩn bị mẫu và có thể sử dụng DNA kém chất lượng hoặc có nồng độ thấp, đồng thời tránh được các lỗi trong quá trình nhân bản khuôn bằng PCR. Phương pháp này cũng sử dụng việc giải trình tự trực tiếp RNA nên loại bỏ được các sai lệch trong nhân bản cDNA. Hiện nay, đã có hai thiết bị giải trình tự theo phương pháp giải trình tự phân tử đơn đó là *Helicos* - Helicos Genetic Analysis System (<http://www.helicosbio.com>) và PacBio *RS SMS* của Pacific BioSciences (<http://www.pacificbiosciences.com>).

Do sự khác nhau về độ dài đoạn đọc, mục đích của từng công nghệ giải trình tự là khác nhau. Với đoạn đọc ngắn và giá thành thấp của Solexa và SOLiD thì hai công nghệ này phù hợp cho giải trình tự toàn bộ hệ gen, trong đó trình tự genome mới có thể lắp ráp và so sánh với trình tự tham chiếu (trình tự genome của loài đang tồn tại). Công nghệ giải trình tự Roche 454 với chuỗi đọc dài (có thể tới 800 bp) cũng có thể sử dụng để nhìn được tổng thể bước đầu về hệ gen và hệ sao chép của loài.

Công nghệ giải trình tự thế hệ mới được sử dụng trong nghiên cứu mã vạch DNA thế hệ mới-next-generation DNA barcoding [81, 161], trong xác định đột biến [19], trong nghiên cứu phân loại và phát sinh loài [41, 164], trong nghiên cứu biến đổi hệ gen và phiên mã ở cơ thể đa bội [72] và trong phát triển chi thị DNA như SNP và SSR [9, 209].

## KẾT LUẬN VÀ KHUYẾN CÁO

Cho đến nay, đã có rất nhiều kỹ thuật được sử dụng để phát triển chi thị DNA phục vụ cho nghiên cứu và chọn lọc ở thực vật. Mỗi kỹ thuật đều có những điểm mạnh và điểm yếu. Việc lựa chọn kỹ thuật chi thị DNA phù hợp và hiệu quả cần dựa vào các đặc điểm: về mức độ đa hình của các chi thị sử dụng cao hay thấp, số lượng locus có thể nhận được trên lần phân tích nhiều hay ít, kiểu di truyền của chi thị trội hay đồng trội, mức độ ổn định của kỹ thuật xây dựng chi thị để có thể lặp lại được, về mức độ đơn giản hay phức tạp của kỹ thuật, số lượng và chất lượng DNA cần thiết cho phân tích, mức độ khó dễ trong sử dụng chi thị, mức độ tự động

hóa và giá thành đầu tư ban đầu và giá thành mỗi lần phân tích. Ngoài ra, cần lưu ý một số đặc điểm khác như: có cần thông tin về trình tự nucleotide không, loại môi sử dụng, có phải sử dụng phóng xạ không và cần nhiều thời gian hay ít. Tuy nhiên, không phải mục đích sử dụng nào cũng cần các kỹ thuật chỉ thị với đầy đủ các đặc điểm đã nêu và cũng không có kỹ thuật chỉ thị nào có đầy đủ các đặc điểm mong muốn. Với mỗi mục đích sử dụng cần một số đặc điểm quan trọng của chỉ thị DNA. Chẳng hạn đối với nghiên cứu đa dạng di truyền, các chỉ thị cần có một số đặc điểm như: phải là chỉ thị đơn giản và nhanh về mặt kỹ thuật, giá thành rẻ, cần lượng mẫu và DNA ít, cho nhiều thông tin, có thể lặp lại trong các nghiên cứu, mức độ sai sót thấp nhất, ghi số liệu dễ và chính xác, có nhiều allel (hàm lượng thông tin cao), không cần biết trước thông tin về genome và cơ thể; còn đối với chọn giống nhờ trợ giúp của chỉ thị phân tử, thì các chỉ thị cần có các đặc điểm là dễ sử dụng, giá thành thấp, cần ít DNA, ổn định về kỹ thuật và có thể lặp lại, cho mức độ đa hình cao, đồng trội và rải đều trong genome. Các tính chất của chỉ thị DNA, những đặc điểm của kỹ thuật phát triển chỉ thị DNA và các đặc tính của genome nghiên cứu là những vấn đề tối quan trọng được khuyến cáo cho việc xem xét và lựa chọn một hay một số kỹ thuật phù hợp cho một nghiên cứu cụ thể và cũng là cơ sở quan trọng để mang lại hiệu quả cao trong việc sử dụng chỉ thị DNA cho nghiên cứu và chọn lọc ở thực vật.

#### TÀI LIỆU THAM KHẢO

1. Abdel-Mawgood A. L., Ahmed M. M. M., Ali B. A., 2006. Application of molecular markers for hybrid maize identification. *J. Food Agr. Environ.*, 4: 176-178.
2. Adams M. D., Kelley J. M., Gocayne J. D., Dubrick M. et al., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 252: 1651-1656.
3. Ajibade S. R., Weeden N. F., Chite S. M., 2000. Inter-simple sequence repeat analysis of genetic relationships in the genus *Vigna*. *Euphytica*, 111: 47-55.
4. Akbari M., Wenzl P., Caig V., Carling J., et al., 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.*, 113: 1409-1420.
5. Akkaya M. S., Bhagwat A. A., Cregan P. B., 1992. Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics*, 132: 1131-1139.
6. 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.
7. Anand P., Rebecca F., Taylor W. J. P., 2000. Transferability of sequence tagged microsatellite site (STMS) primers across four major pulses. *Plant Mol. Biol. Rep.*, 18 (4): 395a.
8. Arens P., Coops H., Jansen J., Vosman B., 1998. Molecular genetic analysis of Black poplar (*Populus nigra* L.) along Dutch rivers. *Mol. Ecol.*, 7: 110-118.
9. Azam S., Thakur V., Ruperao P., Shah T. et al., (2012) Coverage-based consensus calling (CbCC) of short sequence reads and comparison of CbCC results for the identification of SNPs in chickpea (*Cicer arietinum*; Fabaceae), a crop species without a reference genome. *Amer. J. Bot.*, 99: 186-192.
10. Baird E., Cooper-Bland S., Waugh R., DeMaine M., Powell W., 1992. Molecular characterisation of inter and intra-specific somatic hybrids of potato using randomly amplified polymorphic DNA (RAPD) markers. *Mol. Gen. Genet. (MGG)*, 233(3): 469-475.
11. Baldwin B. G., Sanderson M, Porterz J., Wojciechowski M. F. et al., 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, 82: 247-277.
12. Barkley N. A., Dean R. E., Pittman R. N., Wang M. L. et al., 2007. Genetic diversity of cultivated and wild-type peanuts

- evaluated with M13-tailed SSR markers and sequencing. *Genet. Res.*, 89: 93- 06.
13. Phan Thị Bẩy, Lê Thị Bích Thủy, Đào Thị Hạnh, Quách Thị Liên, Lê Thị Muội, Nguyễn Đức Thành, 2005. Đánh giá tính kháng bệnh đạo ôn ở một số dòng lúa dựa vào các chỉ thị phân tử STS và gây bệnh nhân tạo trong nhà kính. *Tạp chí Công nghệ Sinh học*, 3(4): 471-478.
  14. Becker J., Vos P., Kuiper M., Salamini F., Heun M., 1995. Combined mapping of AFLP and RFLP markers in barley. *Mol. Gen. Genet.*, 249: 65-73.
  15. Beckmann J. S., 1988. Oligonucleotide polymorphisms: A new tool for genomic genetics. *Biotechnol.*, 6: 161-164.
  16. Beckmann J. S., Soller M., 1990. Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technol.*, 8: 930-932.
  17. Bentley D. R., 2006. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.*, 16: 545-552.
  18. Bhagyawant S. S., Srivastava N., 2008. Genetic fingerprinting of chickpea (*Cicer arietinum* L.) germplasm using ISSR markers and their relationships. *Afr. J. Biotechnol.*, 7(24): 4428-4431.
  19. Binh T. T., Shiota S., Suzuki R., Matsuda M. et al., 2014. Discovery of novel mutations for clarithromycin resistance in *Helicobacter pylori* by using next-generation sequencing. *J. Antimicrob. Chemother.*, 69: 1796-1803.
  20. Botstein D., White R. L., Skolnick M., Davis R. W., 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Amer. J. Hum. Genet.*, 32: 314-331.
  21. Bowers J. E., Meredith C. P., 1994. Microsatellite length polymorphism within ancient wine grape cultivars (*Vitis vinifera* L.). II Intern Conf. Plant Genome, 24-27. San Diego, CA, USA.
  22. Branco C. J. S., Vieira E. A., Malone G., Kopp M. M. et al., 2007. IRAP and REMAP assessments of genetic similarity in rice. *J. Appl. Genet.*, 48: 107-113.
  23. Bùi Chí Bửu, Nguyễn Thị Lang, 1999. Ứng dụng DNA marker trong đánh giá quỹ gen cây lúa: 1216-1273. Báo cáo khoa học, Hội nghị Công nghệ sinh học toàn quốc, Hà Nội.
  24. Bravo J. P., Hoshino A. A., Lara C. D., Angelici C. M. et al., 2006. Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. *Genet. Mol. Biol.*, 29(3): 516-524.
  25. Brouner S., Murphy R. L., Walling J. G., Przyborowski J., Weeden N. F., 2002. STS marker for comparative mapping in legumes. *J. Amer. Soc. Hort. Sci.*, 127(4): 616-622.
  26. Bruns T. D., White T. J., Taylor J. W., 1991. Fungal molecular systematics. *Annu. Rev. Ecol. Syst.*, 22: 525-564.
  27. Caetano-Anolle's G., Bassam B. J., 1993. Amplification fingerprinting using arbitrary oligonucleotide primers. *App. Biochem. Biotechnol.*, 42: 189-200.
  28. Caetano-Anolles G., Bassam B. J., Gresshoff P. M., 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. *Biotechnol.*, 9: 553-557.
  29. Carvalho A., Matos M., Lima-Brito J., Guedes-Pinto H., Benito C., 2005. DNA fingerprint of F1 interspecific hybrids from the *Triticeae* tribe using ISSRs. *Euphytica*. 143(1-2): 93-99.
  30. Chambers G. K., MacAvoy E. S., 2000. Microsatellites: consensus and controversy. *Comp. Biochem. Physiol. Biochem. Mol. Biol.*, 126(4): 455-476.
  31. Chelkowski J., Stepień L., 2001. Molecular markers for leaf rust resistance genes in wheat. *J. Appl. Genet.*, 42:117-126.

32. Chen J., Tauer C. G., Huang Y., 2002. Paternal chloroplast inheritance patterns in pine hybrids detected with trnL-trnF intergenic region polymorphism. *Theor. Appl. Genet.*, 104: 1307-1311.
33. Cheng H. Y., Yang W. C., Hsiao J. Y., 2001. Genetic diversity and relationship among peach cultivars based on random amplified microsatellite polymorphism (RAMP). *Bot. Bull. Acad. Sin.*, 42: 201-206.
34. Chen S., Chen G., Chen H., Wei Y. et al., 2012. Mapping stripe rust resistance gene *YrSph* derived from *Triticum sphaerococcum* Perc. with SSR, SRAP, and TRAP markers. *Euphytica*, 185(1): 19-26.
35. Ching A. D. A., Caldwell K. S., Jung M., Dolan M. et al., 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet.*, 3: 19.
36. Clausen A. M., Spooner D. M., 1998. Molecular support for the hybrid origin of the wild potato species *Solanum × rechei*. *Crop Sci.*, 38:858-865.
37. Nguyễn Thị Phương Đoài, Khuất Hữu Trung, Nguyễn Thúy Diệp, Hà Minh Loan, Trần Danh Sửu, Đặng Trọng Lương, 2010. Nghiên cứu đa dạng di truyền tập đoàn lúa nương bản địa Việt Nam bằng chỉ thị phân tử SSR. *Tạp chí Công nghệ sinh học*, 4: 381-287.
38. Dietrich W., Katz H., Lincoln S. E., Shin H. S. et al., 1992. A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics*, 131: 423-447.
39. Dintinger J. A., Salgon Sand Reynaud B., 2014. QTL mapping of a partial resistance to the corn delphacid-transmitted viruses in Lepidopteran-resistant maize line Mp705. *Plant Breed.*, 133(1): 19-27.
40. Dissanayaka S., Kottearachchi N.S., Weerasena J., Peiris M., 2014. Development of a CAPS marker for the *badh2.7* allele in Sri Lankan fragrant rice (*Oryza sativa*). *Plant Breed.*, 133(5):560-565.
41. Duarte J. M., Wall K., Edger P., Landherr L. L. et al., 2010. Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.*, 10: 61.
42. Dubreuil P., Dufour P., Krejci E., Causse M. et al., 1996. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. *Crop Sci.*, 36: 790-799.
43. Durand J., Garnier L., Dajoz I., Mousset S., Veuille M., 2000. Gene flow in a facultative apomictic Poacea, the savanna grass *Hyparrhenia diplandra*. *Genetics*, 156: 823-831.
44. Duran C. N., Appleby T., Clark D., Wood M. et al., 2009. AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. *Nucleic Acids Res.*, 37: 951-953.
45. Echt C. S., DeVerno L. L., Anzidei M. A., Vendramin G. G., 1998. Chloroplast microsatellites reveal population diversity in red pine, *Pinus resinosa* Ait. *Mol. Ecol.*, 7: 307-316.
46. Edwards D., Batley J., 2010. Plant genome sequencing: applications for plant improvement. *Plant Biotech. J.*, 8: 2-9.
47. Esselink G. D., Smulders M. J. M., Vosman B., 2003. Identification of cut rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite site markers. *Theor. Appl. Genet.*, 106: 277-286.
48. Fang D., Krueger R. R., Roose M. L., 1998. Phylogenetic Relationships among Selected *Citrus* Germplasm Accessions Revealed by Inter-simple Sequence Repeat (ISSR) Markers. *J. Amer. Soc. Hort. Sci.*, 123(4): 612-617.
49. Fang D. Q., Roose M. L., 1997. Identification of closely related citrus cultivars with inter-simple sequence repeat markers. *Theor. Appl. Genet.*, 95(3): 408-417.

50. Flavell A. J., Knox M. R., Pearce S. R., Ellis T. H. N., 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J.*, 16: 643-660.
51. Flandez-Galvez H., Ford R., Pang E.C.K., Taylor P. W. J., 2003. An intraspecific linkage map of the chickpea (*Cicer arietinum* L.) genome based on sequence tagged microsatellite site and resistance gene analog markers. *Theor. Appl. Genet.*, 106(8): 1447-1456.
52. Fouly H. M., Wilkinson H. T., Domier L. L., 1996. Use of random amplified polymorphic DNA (RAPD) for identification of *Gaeumannomyces* species. *Soil Biol. Biochem.*, 28(6): 703-710.
53. Frascaroli E., Schrag T. A., Melchinger A. E., 2013. Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR and SNP markers reveals ascertainment bias for a subset of SNPs. *Theor. Appl. Genet.*, 126: 133-141.
54. Graner A., Jahoor A., Schondelmaier J., Siedler H. et al., 1991. Construction of an RFLP map of barley. *Theor. Appl. Genet.*, 83(2): 250-256.
55. Garcia V. F., Olmstead R. G., 2003. Phylogenetics of tribe Antocercideae (Solanaceae) based on *ndhF* and *trnL/F* sequence data. *Syst. Bot.*, 28(3): 609-615.
56. Grzebelus D., 2006. Transposon insertion polymorphism as a new source of molecular markers. *J. Fruit Ornam. Plant Res.*, 14: 21-29.
57. Gu W. K., Weeden N. F., Yu J., Wallace D. H., 1995. Large-scale, cost-effective screening of PCR products in marker-assisted selection applications. *Theor. Appl. Genet.*, 91: 465-470.
58. Guasmi F., Elfalleh W., Hannachi H., Fères K. et al., 2012. The use of ISSR and RAPD markers for genetic diversity among South Tunisian barley. *ISRN Agronomy*, vol. 2012, Article ID 952196.
59. Gulsen O., Karagul S., Abak K., 2007. Diversity and relationships among Turkish okra germplasm by SRAP and phenotypic marker polymorphism. *Biologia, Bratislava*, 62: 41-45.
60. Gupta M., Chyi Y. S., Romero-Severson J., Owen J. L., 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. *Theor. Appl. Genet.*, 89: 998-1006.
61. Guryev V., Berezikov E., Cuppen E., 2005. CASCAD: a database of annotated candidate single nucleotide polymorphisms associated with expressed sequences. *BMC Genomics*, 6: 10. doi:10.1186/1471-2164-6-10.
62. Ha B. K., Hussey R. S., Boerma H. R., 2007. Development of SNP assays for marker-assisted selection of two southern root-knot nematode resistance QTL in soybean. *Crop Sci.*, 47(2): S73-S82.
63. Hafez E. E., Ghany A. G. A. A., Zakil E. A., 2006. LTR-retrotransposon-based molecular markers in cultivated Egyptian cottons *G. barbadense* L. *Afr. J. Biotechnol.*, 5: 1200-1204.
64. Hamada H., Petrino M. G., Kakunaga T., 1982. A novel repeat element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci., USA* 79:6465-6469.
65. Hansen O. K., Kjær E. D., Vendramin G. G., 2005. Chloroplast microsatellite variation in *Abies nordmanniana* and simulation of causes for low differentiation among populations. *Tree Genetics & Genomes*, 1: 116-123.
66. Hantula J., Dusabenygasani M., Hamelin R.C., 1996. Random amplified microsatellites (RAMS)-a novel method for characterizing genetic variation within fungi. *Eur. J. For. Path.*, 26: 15-166.
67. He C., Poysa V., Yu K., 2003. Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. *Theor. Appl. Genet.*, 106: 363-373.

68. Heun M., Schafer-Pregl R., Klawan D., Castagna R. et al., 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. *Science*, 278: 1312-1314.
69. Hill M., Witsenboer H., Zabeau M., Vos P. et al., 1996. PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. *Theor. Appl. Genet.*, 93: 1202-1210.
70. Huang J., Sun S.M., 2000. Genetic diversity and relationships of sweet potato and its wild relatives in *Ipomoea* series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. *Theor. Appl. Genet.*, 100: 1050-1060.
71. Vũ Thị Bích Huyền, Lê Thị Bích Thủy, Nguyễn Anh Dũng, Hoàng Bá Tiên, Nguyễn Đức Thành, 2013. Đánh giá đa dạng di truyền một số giống lúa bằng kỹ thuật SSR phục vụ chọn cặp lai tạo giống chịu hạn. *TẠP CHÍ SINH HỌC*, 35(1): 80-91.
72. Ilut D. C., Coate J. E., Luciano A. K., Owens T. G. et al., 2012 A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. *Amer. J. Bot.*, 99: 383-396.
73. Jaccoud D., Peng K., Feinstein D., Kilian A., 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Res.*, 29(4): e25.
74. Jakse J., Martin W., McCallum J., Havey M., 2005. Single nucleotide polymorphisms, InDel, and simple sequence repeats for onion cultivar identification. *J. Amer. Hort. Sci.*, 130(6): 912-917.
75. Jakob S. S., Ihlow A., Blattner F. R., 2007. Combined ecological niche modeling and molecular phylogeography history of *Hordeum marinum* (Poaceae)-niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Mol. Ecol.*, 16: 1713-1727.
76. Jehan T., Lakhanpaul S., 2006. Single nucleotide polymorphism (SNP)- methods and applications in plant genetics: A review. *Indian J. Biotech.*, 5: 435-459.
77. Jing R., Vershinin A., Grzebyta J., Shaw P. et al., 2010. The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol. Biol.*, 10:44.
78. Jordan S. A., Humphries P., 1994. Single nucleotide polymorphism in exon 2 of the BCP gene on 7q31-q35. *Hum. Mol. Genet.*, 3: 1915.
79. Joshi S. P., Gupta V. S., Aggarwal R. K., Ranjekar P. K. et al., 2000. Genetic diversity and phylogenetic relationship as revealed by inter-simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. *Theor. Appl. Genet.*, 100: 1311-1320.
80. Kallendar R., Grob T., Regina M., Suoniemi A., Schulman A., 1999. IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. *Theor. Appl. Genet.*, 98: 704-711.
81. Kane N., Sveinsson S., Dempewolf H., Ang J. Y. Y. et al., 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. *Amer. J. Bot.*, 99: 320-329.
82. Karp A., Isaac P. G., Ingram G. S., 1998. *Molecular Tools for Screening Biodiversity: Plants and Animals*. Chapman & Hall, Thompson Sci., London.
83. Kaur S., Cogan N. O. I., Forster J. W., Paull J. G., 2014 Assessment of genetic diversity in Faba bean based on single nucleotide polymorphism. *Diversity*, 6: 88-101; doi:10.3390/d6010088.
84. Kawa P. G., Devarumath R. M., Nerka Y., 2009. Use of RAPD marker for assessment of genetic diversity in sugarcane cultivars. *Indian J. Biotechnol.*, 8: 67-81.

85. Khan A., Awan S., Sadia B., Rana R. M., Khan I. A., 2010. Genetic diversity study among coloured cotton genotypes by using RAPD markers. *Pak. J. Bot.*, 42(1): 71-77.
86. Kijas J. M. H., Fowler J. C. S., Thomas M. R., 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. *Genome*, 38(2): 349-355.
87. Kim H. M., Oh S. H., Bhandari G. S., Kim C. S., Park C. W., 2014. DNA barcoding of Orchidaceae in Korea. *Mol. Ecol. Resources*, 14(3): 499-507.
88. Kim K. S., Bellendir S., Hudson K. A. et al., 2010 Fine mapping the soybean aphid resistance gene *Rag1* in soybean. *Theor. Appl. Genet.*, 120(5): 1063-1107.
89. Kim S. M., Sohn J. K., 2005. Identification of a rice gene (*Bph 1*) conferring resistance to brown planthopper (*Nilaparvata lugens* Stal) using STS markers. *Mol. Cells*, 20(1): 30-34.
90. Kim S. K., Crawford D. J., Esselman E. J., 1999. ITS sequences and phylogenetics relationships in *Bidens* and *Coreopsis* (Asteraceae). *Syst. Bot.*, 24: 480-491.
91. Kocyan A., Qui Y. L., Endress P. K., Conti E., 2004. A phylogenetics analysis of Apostasioideae (Orchidaceae) based on ITS, *trnL-F* and *matK* sequences. *Plant Syst. Evol.*, DOI 10.1007/s00606-004-0133-3.
92. Koes R., Souer E., van Houwelingen A., Mur L. et al., 1995. Targeted gene inactivation in petunia by PCR-based selection of transposon insertion mutants. *Proc. Natl. Acad. Sci.*, 92: 8149-8153.
93. Konieczny A., Ausubel F. M., 1993. Procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.*, 4: 403-410.
94. Kwok P. Y., Chen X., 2003. Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.*, 5: 43-60.
95. Kwon S. J., Smýkal P., Hu J., Wang M. et al., 2013. User-friendly markers linked to Fusarium wilt race 1 resistance *Fw* gene for marker-assisted selection in pea. *Plant Breed.*, 132(6): 642-648.
96. Lander E. S., Botstein D., 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, 121: 185-199.
97. Landegren U., Kaiser R., Sanders J., Hood L., 1988. DNA diagnostics. Molecular techniques and automation. *Science*, 241: 1077-1080.
98. Nguyen Thi Lang, Bui Chi Buu, 2008. Fine mapping for drought tolerance in rice (*Oryza sativa* L.). *Omonrice*, 16: 9-15.
99. Nguyen Thi Lang, Nguyen Van Tao, Bui Chi Buu, 2011. Marker-assisted backcrossing (MAB) for rice submergence tolerance in Mekong Delta. *Omonrice*, 18: 11-21
100. Lee G. A., Koh H. J., Chung H. K., Dixit A. et al., 2009. Development of SNP-based CAPS and dCAPS markers in eight different genes involved in starch biosynthesis in rice. *Mol. Breed.*, 25(1): 93-101.
101. Li A., Ge S., 2001. Genetic Variation and Clonal Diversity of *Psammochloa villosa* (Poaceae) Detected by ISSR Markers. *Ann. Bot.*, 87: 585-590.
102. Li G., Quiros C. F., 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in Brassica. *Theor. Appl. Genet.*, 103: 455-546.
103. Li X., Ding X., Chu B., Zhou Q. et al., 2008. Genetic diversity analysis and conservation of the endangered Chinese endemic herb *Dendrobium officinale* Kimura et Migo (Orchidaceae) based on AFLP. *Genetica*, 133: 159-166.
104. Nguyễn Thị Kim Liên, Nông Văn Hải, Nguyễn Đức Thành, 2003. Kết quả bước đầu của việc ứng dụng chỉ thị SSR trong chọn dòng chịu hạn ở lúa cạn. Báo cáo Khoa học, Hội nghị Công nghệ sinh học toàn quốc, Hà Nội, 898-901.

105. Lin X. C, Lou Y., Liu J., Peng J. S. et al., 2010. Crossbreeding of *Phyllostachys* species (Poaceae) and identification of their hybrids using ISSR markers. *Genet. Mol. Res.*, 9(3): 1398-1404.
106. Lin Z., He D., Zhang X., Nie Y. et al., 2005. Linkage map construction and mapping QTL for Cotton fibre quality using SRAP, SSR, and RAPD. *Plant Breed.*, 124: 180-187.
107. Litt M., Luty J. A., 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Amer. J. Hum. Genet.*, 44(3): 397-401.
108. Lu X., Liu L., Zhao L., Song X., Zhu X., 2009. Cultivar identification and genetic diversity analysis of broccoli and its related species with RAPD and ISSR markers. *Sci. Hort.*, 122(4): 645-648.
109. Trần Thị Lương, Lưu Minh Cúc, Nguyễn Đức Thành, 2013. Phân tích quan hệ di truyền một số giống lúa đặc sản, chất lượng trồng phổ biến ở Việt Nam bằng chỉ thị phân tử SSR. *TẠP CHÍ SINH HỌC*, 35(3): 348-356.
110. Ma J. Q., Yao M. Z., Ma C. L., Wang X. C. et al., 2014. Construction of a SSR-based genetic map and identification of QTLs for Catechins content in tea plant (*Camellia sinensis*). *PLoS ONE*, 9(3): e93131. doi:10.1371/journal.pone.0093131.
111. McClintock B., 1950. The origin and behavior of mutable loci in maize. *Genetics*, 36: 344-355.
112. Margulies M., Egholm M., Altman W.E., Attiya S. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437: 376-380.
113. Mardis E. R., 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.*, 24: 133-141.
114. Martin C., Uberhuaga E., Pérez C., 2002. Application of RAPD markers in the characterisation of *Chrysanthemum* varieties and the assessment of somaclonal variation. *Euphytica*, 127(2): 247-253.
115. McDonald D. B., Potts W. K., 1997. Microsatellite DNA as a genetic marker at several scales. In *Avian Molecular Evolution and Systematics* (D. Mindell, ed.). Academic Press, New York. pp. 29-49.
116. Mengoni A., Barabesi C., Gonnelli C., Galardi F. et al., 2001. Genetic diversity of heavy metal-tolerant populations in *Silene paradoxa* L. (Caryophyllaceae): A chloroplast microsatellite analysis. *Mol. Ecol.*, 10:1909-1916.
117. Meti N., Samal K. C., Bastia D. N., Rout G. R. 2013. Genetic diversity analysis in aromatic rice genotypes using microsatellite based simple sequence repeats (SSR) marker. *Afr. J. Biotechnol.*, 12(27): 4238-4250.
118. Michaels S. D., Amasino R. M. A., 1998. A robust method for detecting single nucleotide changes as polymorphic markers by PCR. *Plant J.*, 14: 381-385,
119. Miller J. C., Tanksley S. D., 1990. RFLP analysis of phylogenetics relationships and genetic variation in the genus *Lycopersicon*. *Theor. Appl. Genet.*, 80: 437-448.
120. Mishra P. K., Fox R. T. V., Culham A., 2003. Inter-simple sequence repeat and aggressiveness analyses revealed high genetic diversity, recombination and longrange dispersal in *Fusarium culmorum*. School of Plant Sciences, The University of Reading, Whiteknights, Reading RG6 6AS, UK, 2003 Association of Applied Biologists.
121. Morand M. E., Brachet S., Rossignol P., Dufour J., Frascaria-Lacoste N., 2002. A generalized heterozygote deficiency assessed with microsatellites in French common ash populations. *Mol. Ecol.*, 11:377-385.
122. Morgante M., Vogel J., 1994. Compound microsatellite primers for the detection of



- genetic polymorphisms. US patent application no. 08/326456.
123. Mullis K. B., Faloona F., 1987. Specific synthesis of DNA in vitro via polymerase chain reaction. *Methods Enzymol.*, 155:350-355.
  124. Nagaoka T., Ogihara Y., 1997. Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. *Theor. Appl. Genet.*, 94: 597-602.
  125. Naik S., Gill K. S., Prakasa Rao V. S., Gupta V. S. et al., 1998. Identification of a STS marker linked to the *Aegilops speltoides*-derived leaf rust resistance gene *Lr28* in wheat *Theor. Appl. Genet.*, 97(4):535-540.
  126. Nair A. S., Teo C. H., Schwarzacher T., Heslop Harrison P., 2005. Genome classification of banana cultivars from South India using IRAP markers. *Euphytica*, 144: 285-290.
  127. Nakamura Y., Leppert M., O'Connell P., Wolff R. et al., 1987. Variable number tandem repeat (VNTR) markers for human gene mapping. *Science*, 235: 1616-1622.
  128. Neff M. M., Neff J. D., Chory J., Pepper A. E., 1998. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant J.*, 14(3): 387-92.
  129. Nicolosi E., Deng Z.N., Gentile A., La Malfa S. et al., 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theor. Appl. Genet.*, 100:1155-1166.
  130. Noguchi J., Hong D. Y., Grant W. F., 2004. The historical evolutionary development of *Hemerocallis middendorfii* (Hemerocallidaceae) revealed by non-coding regions in chloroplast DNA. *Plant Syst. Evol.*, 247: 1-22.
  131. Obeed R. S., Harhash M. M., Abdel-Mawgood A.L., 2008. Fruit properties and genetic diversity of five ber (*Ziziphus mauritiana* Lamk) cultivars. *Pak. J. Biol. Sci.*, 11:888-893.
  132. Olmstead R. G., Michaels H. J., Scott K. M., Palmar J. D., 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcl*. *Ann Missouri Bot. Gard.*, 79:249-265.
  133. Olmstead R. G., Palmer J. D., 1997. Implication for phylogeny, classification, and biogeography of *Solanum* from cpDNA restriction sites variation. *Syst. Bot.*, 22(1):19-29.
  134. Olsen M., Hood L., Cantor C., Botstein D., 1989. A common language for physical mapping of the human genome. *Science* 245:1434-1435.
  135. Olufowote J. O., Xu Y., Chen X., Goto M. et al., 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome*, 40(3): 370-378.
  136. Orita M., Iwahana H., Kanazawa H., Hayashi K., Sekiya T., 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphism. *Proc. Natl. Acad. Sci., USA* 86:2766-2770.
  137. Paran I., Michelmore R. W., 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theor. Appl. Genet.*, 85:985-999.
  138. Pearce S. R., Knox M., Ellis T. H. N., Flavell A. J., Kumar A., 2000. Pea Ty1-copia group of retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. *Mol. Gen. Genet.*, 263:898-907.
  139. Perry D. J., Bousquet J., 2001. Genetic diversity and mating system of post-fire and post-harvest black spruce: an investigation using codominant sequence-

- tagged-site (STS) markers. *Canadian J. Forest Res.*, 31(1): 32-40.
140. Petersen G., Seberg O., 1996. ITS regions highly conserved in cultivated barleys. *Euphytica*, 90: 233-234.
  141. Pharmawati M., Yan G., Sedgley R., Finnegan P. M., 2004. Chloroplast DNA inheritance and variation in *Leucadendron* species (Proteaceae) as revealed by PCR-RFLP. *Theor. Appl. Genet.*, 109: 1694-1701.
  142. Provan J., Corbett G., McNicol J.W., Powell W., 1997. Chloroplast variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. *Genome*, 40: 104-110.
  143. Provan J., Russell J. R., Booth A., Powell W., 1999. Polymorphic simple sequence repeat primers for systematic and population studies in the genus *Hordeum*. *Mol. Ecol.*, 8: 505-511.
  144. Queen R. A., Gribbon B. M., James C., Jack P., Flavell A. J., 2004. Retrotransposon based molecular markers for linkage and genetic diversity analysis in wheat. *Mol. Genet. Genom.*, 271: 91-97.
  145. Qian W., Ge S., Hong D. Y., 2001. Genetic variation within and among populations of a wild rice *Oryza granulata* from China detected by RAPD and ISSR markers. *Theor. Appl. Genet.*, 102(2-3): 440-449.
  146. Raman H., Dalton-Morgan J., Diffey S., Raman R., Alamery S. et al., 2014. SNP markers-based map construction and genome-wide linkage analysis in *Brassica napus*. *Plant Biotech. J.*, 12 (7): 851-860.
  147. Ratnaparkhe M. B., Tekeoglu M., Muehlbauer F. J., 1998. Inter-simple-sequence-repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene clusters. *Theor. Appl. Genet.*, 97(4): 515-519.
  148. Raucher J. T., Doyle J. J., Brown H. D., 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Mol. Ecol.*, 11: 2691-2702.
  149. Reddy P. M., Sarla N., Siddiq E. A., 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. *Euphytica*, 128-2: 9-17.
  150. Ritland K., Clegg M. T., 1987. Evolutionary and DNA sequences. *Amer. Naturalist.*, 30: S74-S100.
  151. Ritland C. E., Ritland R. K., Straus N. A., 1993. Variation in the ribosomal internal transcribed spacers (ITS1 and ITS2) among eight taxa of the *Mimulus guttatus* species complex. *Mol. Biol. Evol.*, 10: 1273-1288.
  152. Rokkan V. M., Xu Y. S., Kankila J., Kuusela A. et al., 1994 Identification of somatic hybrids of dihaploid *Solanum tuberosum* lines and *S. brevidens* by species specific RAPD patterns and assessment of disease resistance of the hybrids. *Euphytica*, 80(3): 207-217.
  153. Rossi I., Bartolacci B., Potenza L., Bertini L. et al., 2000. Identification of white truffle species using RAPD markers. *Plant and Soil*, 219(1-2): 127-133.
  154. Rothberg J. M., Hinze W., Rearick T. M., Schultz J. et al., 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475: 348-352.
  155. Sachidanandam R., Weissman D., Schmidt S. C., Kakol J. M. et al., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409:928-933.
  156. Salimath S. S., de Oliveira A.C., Godwin I.D., Bennetzen J.L., 1995 Assessment of genome origins and genetic diversity in the genus *Eleusine* with DNA markers. *Genome*, 38: 757-763.
  157. Sanchez de la Hoz M. P., Davila J. P., Loarce Y., Ferrer E., 1996. Simple sequence repeat primers used in polymerase chain reaction amplifications

- to study genetic diversity in barley. *Genome*, 39:112-117.
158. Shao K., Ding W., Wang F., Li H. et al., 2011. Emulsion PCR: A High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection. *PLoS ONE* 6(9): e24910.
159. Sharma S. K., Knox M. R., Ellis T. H. N., 1996. AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theor. Appl. Genet.*, 93: 751-758.
160. Shendure J., Porreca G. J., Reppas N. B., Lin X. et al., 2005. Accurate multiplex colony sequencing of an evolved bacterial genome. *Science*, 309: 1728-1732.
161. Shokralla S., Gibson J. F., Nikbaht H., Janzen D. H. et al., 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol. Ecol. Resources*, 14: 892-901.
162. Slabaugh M. B., Huestis G. M., Leonard J., Holloway J. L. et al., 1997. Sequence-based genetic markers for genes and gene families: single-strand conformational polymorphisms for the fatty acid synthesis genes of *Cuphea*. *Theor. Appl. Genet.*, 94:400-408.
163. Soleimani M. H., Talebi M., Sayed-Tabatabaei B. E., 2012. Use of SRAP markers to assess genetic diversity and population structure of wild, cultivated, and ornamental pomegranates (*Punica granatum* L.) in different regions of Iran. *Plant Syst. Evol.*, 298(6): 1141-1149.
164. Straub S.C.K., Parks M., Weitemier K., Fishbein M. et al., 2012. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. *Amer. J. Bot.*, 99(2): 349-364.
165. Suh Y., Vigg J., 2005. SNP discovery in associating genetic variation with human disease phenotypes. *Mutat. Res.*, 573: 41-53.
166. Sun Z., Wang Z., Tu J., Zhang J. et al., 2007. An ultradense genetic recombination map for *Brassica napus*, consisting of 13551 SRAP markers. *Theor. Appl. Genet.*, 114(8):1305-1317.
167. Taberlet P., Gielly L., Pautou G., bouvet J., 1991. Universal Primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.*, 17: 1105-1109.
168. Tatikonda L., Wani S. P., Kannan S., Beerelli N. et al., 2009. AFLP-based molecular characterization of an elite germplasm collection of *Jatropha curcas* L.: A biofuel plant. *Plant Sci.*, 176: 505-513.
169. Tautz D., Renz M., 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.*, 12(10):4127-4138.
170. Telenius H., Carter N. P., Bebb C. E., Nordenskjold M. et al., 1992. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, 13: 718-725.
171. Nguyễn Đức Thành, Henry Nguyễn, 1999. Nghiên cứu đa dạng phân tử ở lúa bằng kỹ thuật đa hình các chuỗi lặp lại đơn giản (SSR). *TẠP CHÍ SINH HỌC*, 21(1b): 107-112.
172. N. D. Thanh, H. G. Zheng, N. V. Dong, L. N. Trinh, M. L. Ali & H. T. Nguyen. 1999. Genetics variation in root morphology and microsatellite DNA loci in upland rice (*Oryza sativa* L.) from Vietnam. *Euphytica*, 105: 43-51.
173. Nguyen Duc Thanh, Nguyen Thi Kim Lien, Pham Quang Chung, Tran Quoc Trong, Le Thi Bich Thuy, and Henry Nguyễn. 2006. Mapping QTLs associated with root traits related to drought resistance in Vietnamese upland rice. *Asean Journal on Sci. & Tech. for Development (AJSTD)*, 23(4): 323-332.
174. Nguyễn Đức Thành, Nguyễn Thuý Hạnh, Trần Quốc Trọng. 2007. Kết quả nghiên cứu một số chuỗi gen lục lạp trong nghiên cứu đa dạng di truyền và xuất xứ cây lâm

- nghiệp Tạp chí Công nghệ sinh học, 5(1): 77-83.
175. Nguyễn Đức Thành, Đặng Thị Minh Lụa, Nguyễn Văn Phương, Lê Thị Bích Thủy, 2008. Phân tích đa dạng di truyền các dòng lúa Tú Lệ và một số giống nếp đặc sản dựa vào các gen và chỉ thị liên quan đến chất lượng hạt gạo. TẠP CHÍ SINH HỌC, 30(2): 153- 159.
176. Nguyen Duc Thanh, Nguyen Hoang Tinh, 2009. Application of molecular markers for the study of genetic diversity of forest trees and sub-tropical fruit species. Adv. in National Science, 3: 373-382.
177. Nguyễn Đức Thành, Nguyễn Văn Phương, Nguyễn Hoàng Nghĩa. 2009. Đa dạng di truyền của loài Sao mạng (*Hopea reticulata* Tardicu) dựa trên phân tích một số chuỗi ADN lục lạp và chỉ thị RAPD. Tạp chí Công nghệ sinh học, 7(2): 203-210.
178. Nguyen Duc Thanh, Henry Nguyen. 2000. Use of AFLP for the study of genetic diversity in upland rice. Adv. in Natural Sciences, 1: 107-114.
179. Nguyễn Đức Thành, Phan Duy Toàn, Nguyễn Hoàng Anh, Henry T. Nguyen. 2000. Ứng dụng chỉ thị RAPD và STS trong nghiên cứu đa dạng di truyền và chọn giống lúa. Báo cáo Hội nghị Sinh học Quốc gia. Những vấn đề nghiên cứu cơ bản trong sinh học, Hà Nội, 7-8/8/2000, 149- 152.
180. Nguyen Duc Thanh, Le Thi Bích Thủy, Nguyen Hoang Nghia, 2012. Genetic diversity of *Azelia xylocarpa* (Kurz) Craib in Vietnam based on analyses of chloroplast markers and random amplified polymorphic DNA (RAPD). Afr. J. Biotechnol., 1(80): 14529-14535.
181. Thomas M. R., Cain P., Scott N. S., 1994. DNA typing of grapevines: a universal methodology and database for describing cultivars and evaluating genetic relatedness. Plant Mol. Biol., 25: 939-949.
182. Thomas M. R., Scott N. S., 1993. Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence tagged sites (STSS). Theor. Appl. Genet., 86: 985-990.
183. Lê Thị Bích Thủy, Nguyễn Văn Trữ, Nguyễn Đức Thành, Hồ Hữu Nhị, 2013. Phân tích đa dạng di truyền một số giống mía bằng chỉ thị SSR và xác định chỉ thị phân tử liên quan đến tính kháng bệnh than. Báo cáo khoa học Hội nghị CNSH toàn quốc 2013, 27/9/2013: 1084-1088.
184. Tippery N. P., Les D. H., 2008. Phylogeneic analysis of the internal transcribed spacer in Menyanthaceae using predicted secondary structure. Mol. Phylogenet. Evol., 49: 528-537.
185. Tivang J., Skroch P. W., Nienhuis J., 1996. Randomly amplified polymorphic DNA (RAPD) variation among and within Artichoke (*Cynara scolymus* L.) cultivars and breeding Populations. J. Amer. Soc. Hort. Sci., 121(5): 783-788.
186. Travis S., Maschinski J., Keim P., 1996. An analysis of genetic variation in *Astragalus cremnophylax* var. *cremnophylax*, a critically endangered plant, using AFLP markers. Mol. Ecol., 5: 735-745.
187. Triglia T., Peterson M. G., Kemp D. J., 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res., 16:8180-8186.
188. Twyman R. M., 2004. SNP discovery and typing technologies for pharmacogenomics. Curr. Top. Med. Chem., 4: 1423-1431.
189. van den Broeck D., Maes T., Sauer M., Zethof J. et al., 1998. Transposon Display identifies individual transposable elements in high copy number lines. Plant J., 13: 121-129.
190. Walker G. T., Little M. C., Nadeau J. G., Shank D. D., 1992. Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. Proc. Natl. Acad. Sci., 89: 392-396.

191. Wang D. G., Fan J. B., Siao C. J., Berno A. et al., 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 280: 1077-1082.
192. Wang Q. D., Zhang T., Wang J. B., 2007. Phylogenetic relationships and hybrid origin of *Potamogeton* species (*Potamogetonaceae*) distributed in China: insights from the nuclear ribosomal internal transcribed spacer sequences (ITS). *Plant Syst. Evol.*, 267(1-4): 65-78.
193. Waugh R., McLean K., Flavell A. J., Pearce S. R. et al., 1997. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (SSAP). *Mol. Gen. Genet.*, 253: 687-694.
194. Waugh R., Bonar N., Baird E., Thomas B., Graner A., Hayes P., Powell W., 1997. Homology of AFLP products in three mapping populations of barley. *Mol. Gen. Genet.*, 255: 311-321.
195. Welsh J., McClelland M., 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.*, 18: 7213-7218.
196. Winfield M. O., Arnold G. M., Cooper F. et al., 1998. A study of genetic diversity in *Populus nigra* subsp. *betulifolia* in the Upper Severn area of the UK using AFLP markers. *Mol. Ecol.*, 7: 3-10.
197. Williams J. G. K., Kubelik A. R., Livak K. J. et al., 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.*, 18: 6531-6535.
198. Winter P., Pfaff T., Udupa S. M., Hüttel B. et al., 1999. Characterization and mapping of sequence-tagged microsatellite sites in the chickpea (*Cicer arietinum* L.) genome. *Mol. Gen. Genet.*, 262(1): 90-101.
199. Wolff K., Morgan-Richards M., 1998. PCR markers distinguish *Plantago major* subspecies. *Theor. Appl. Genet.*, 96: 282-286.
200. Vos P., Hogers R., Bleeker M., Reijans M. et al., 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.*, 23: 4407-4414.
201. Wu K. S., Jones R., Danneberger L., Scolnik P., 1994. Detection of microsatellite polymorphisms without cloning. *Nucleic Acids Res.*, 22: 3257-3258.
202. Wu K. S., Tanksley S. D., 1993. Abundance, polymorphism and genetic mapping of microsatellites in rice. *Mol. Gen. Genet.*, 241, 225-235.
203. Wu X., Li Y., Shi Y., Song Y. et al., 2014. Fine genetic characterization of elite maize germplasm using high-throughput SNP genotyping. *Theor. Appl. Genet.*, 127(3): 621-631.
204. Xie J., Wehner T. C., Conkling M. A., 2002. PCR-based single stranded conformation polymorphism (SSCP) analysis to clone nine aquaporin genes in cucumber. *J. Amer. Hort. Sci.*, 127(6): 925-930.
205. Xu D. H., Ban T., 2004. Phylogenetic and evolutionary relationships between *Elymus humidus* and other *Elymus* species based on sequencing of noncoding regions of cpDNA and AFLP of nuclear DNA. *Theor. Appl. Genet.*, 108: 1443-1448.
206. Yashitola J., Thirumurugan T., Sundaram R.M., Naseerullah M. K. et al., 2002. Assessment of purity of rice hybrids using microsatellite and STS markers. *Crop Sci.*, 42: 1369-1373.
207. Yao H., Song J., Liu C., Luo K. et al., 2010. Use of ITS2 Region as the universal DNA barcode for plants and animals. *PLoS ONE* 5(10): e13102. doi:10.1371/journal.pone.0013102
208. Yen A. C., Olmstead R. G., 2000. Molecular systematics of Cyperaceae tribe Cariceae based on two chloroplast DNA regions: *ndhF* and *trnL* intron-intergenic spacer. *Syst. Bot.*, 25(3): 479-494.
209. Zalapa J. E., Cuevas H., Zhu H., Steffan S.

- et al., 2012. Using next-generation sequencing approaches for the isolation of simple sequence repeat (SSR) loci in the plant sciences. *Amer. J. Bot.*, 99: 193-208.
210. Zeid M., Schon C., Link W., 2003. Genetic diversity in recent elite faba bean lines using AFLP markers. *Theor. Appl. Genet.*, 107: 1304-1314.
211. Zhang Z., Kudo T., Nakajima Y., Wang Y., 2001. Clarification of the relationship between the members of the family Thermomonosporaceae on the basis of the rDNA, 16S-23S rRNA internal transcribed spacer and 23S rDNA sequences and chemotaxonomic analysis. *Int. J. Syst. Evol. Microbiol.*, 51: 373-383.
212. Zietkiewicz E., Rafalski A., Labuda D., 1994. Genome fingerprinting by simple sequence repeats (SSR)-anchored PCR amplification. *Genomics*, 20: 176-183.

## DNA MARKER TECHNIQUES IN STUDY AND SELECTION OF PLANT

Nguyen Duc Thanh

Institute of Biotechnology, VAST

### SUMMARY

Since the 1980s, DNA marker techniques have been invented and developed quickly to become the most significant development in the field of molecular biology. The DNA markers have been widely used in study and selection of plants. The DNA techniques have been developed for DNA markers used in studies of genetic diversity, phylogeny, classification, gene tagging and identification; in germplasm and marker-assisted selection. The presence of various types of DNA markers, and differences in their principles, methodologies, and applications require careful consideration in choosing one or more of such methods. No DNA markers are available that fulfill all requirements needed by researchers. Depending on the nature of each study, one can choose among the variety of DNA marker techniques, each of which combines at least some desirable properties. In Vietnam, the use of DNA marker techniques began at the end of 1990s. However, the application was limited with only few techniques such as random amplified DNA polymorphism, microsatellites or simple sequence repeats and amplified fragment length polymorphism. Those techniques were used in the studies of plant genetic diversity, molecular mapping and marker-assisted selection. This review provides an overview on most of the available DNA marker techniques and their utilities in the study and selection of plants with the aim to provide researchers and breeders necessary information for choosing appropriate DNA marker techniques.

*Keywords:* DNA marker techniques, gene identification, genetic diversity, germplasm selection, marker-assisted selection.

*Ngày nhận bài:* 10-5-2014