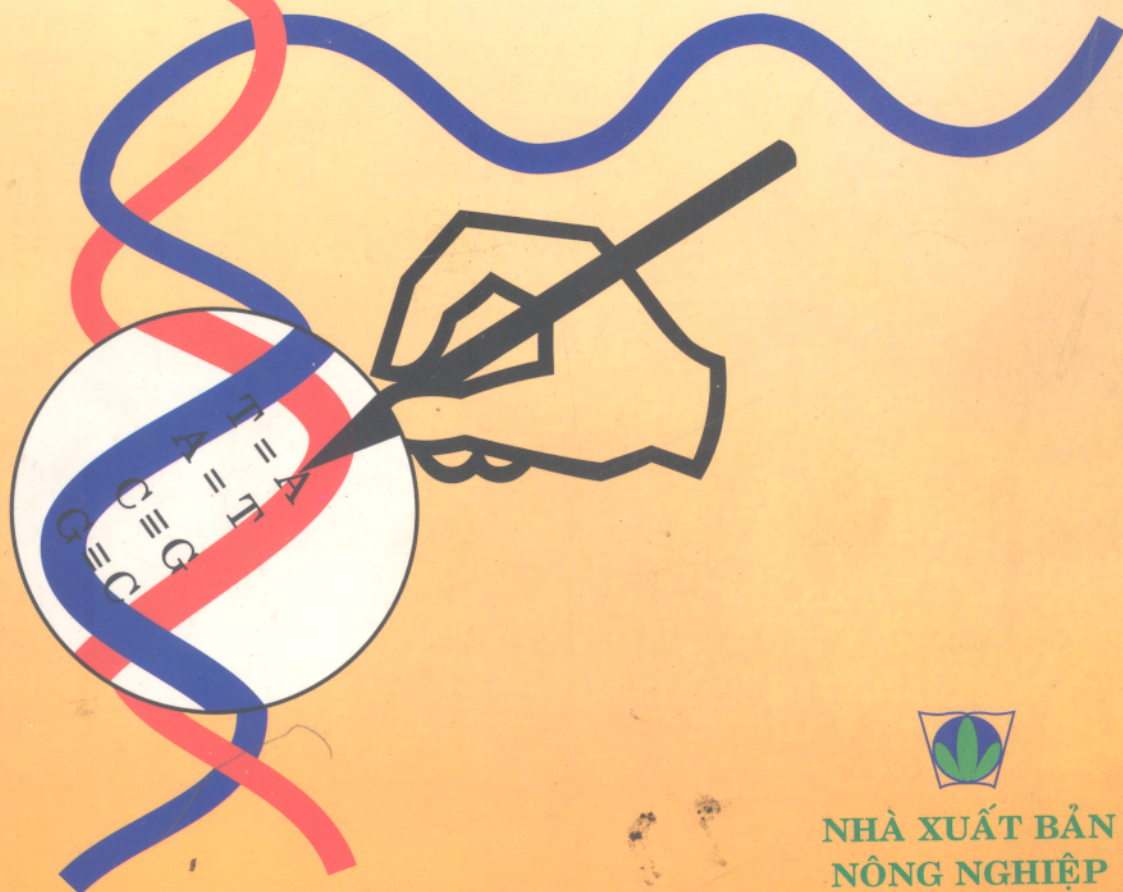


BÙI CHÍ BỬU - NGUYỄN THỊ LANG

DI TRUYỀN PHÂN TỬ

Những nguyên tắc cơ bản
trong chọn giống cây trồng



NHÀ XUẤT BẢN
NÔNG NGHIỆP

BÙI CHÍ BỬU & NGUYỄN THỊ LANG

DI TRUYỀN PHÂN TỬ

**NHỮNG NGUYÊN TẮC CƠ BẢN
TRONG CHỌN GIỐNG CÂY TRỒNG**

**NHÀ XUẤT BẢN NÔNG NGHIỆP
TP. HỒ CHÍ MINH, 1999**

Quyển I

PHÂN TÍCH GENOME

BÙI CHÍ BỬU

Phó Giáo Sư, Phó Tiến Sĩ

Viện Lúa Đồng bằng sông Cửu Long

NGUYỄN THỊ LANG

Phó Tiến Sĩ

Thực tập sinh sau Tiến Sĩ (Post-doct)

Viện Lúa Quốc tế, Philippines

Quyển I

PHÂN TÍCH GENOME

MỞ ĐẦU

Di truyền phân tử là ngành học đang phát triển rất nhanh, trên cơ sở sinh học phân tử, sinh hóa, di truyền, tin học ứng dụng, toán sinh học... Nó là những môn cơ bản để phát triển công nghệ sinh học phục vụ cho nông nghiệp, y học, và những lĩnh vực có liên quan đến sinh học. Quyển sách này chỉ nêu những nguyên tắc cơ bản, có giá trị áp dụng trong cải tiến giống cây trồng. Giáo trình được chia ra làm hai quyển: Quyển I nói về phân tích genome, và Quyển II nói về chuyển nạp gen.

Nếu viết cho đầy đủ nội dung của di truyền phân tử, nó sẽ yêu cầu rất nhiều ngành học có liên quan đến sinh hóa, sinh học phân tử, di truyền vi sinh vật, enzym học phân tử, cấu trúc và chức năng của màng, hoạt động và thể hiện của gen, tương tác ở mức độ phân tử giữa cây chủ và ký sinh... Chúng tôi hi vọng những kiến thức có tính chất nguyên tắc được tóm lược trong quyển sách này sẽ giúp ích phần nào cho các bạn sinh viên, các cán bộ nghiên cứu những tư liệu cần thiết, áp dụng cho việc cải tiến giống cây trồng. Nhưng với nguồn thông tin rất lớn và đa dạng, chắc chắn chúng tôi sẽ còn nhiều thiếu sót, rất mong sự góp ý của bạn đọc.

Nhiều thuật ngữ khoa học có tính chất quốc tế hóa, và thuật ngữ chuyên môn rất khó dịch ra tiếng Việt, chúng tôi tạm sử dụng thuật ngữ tiếng Anh, nhằm mục đích giúp bạn đọc tra cứu sách nước ngoài thuận lợi hơn trong quá trình học tập và nghiên cứu. Giống như lĩnh vực tin học, có nhiều thuật ngữ trong sử dụng câu lệnh phải dùng tiếng Anh, công nghệ sinh học cũng có những yêu cầu như vậy. Rất mong bạn đọc thứ lỗi, khi phát hiện trong sách có quá nhiều thuật ngữ chuyên môn bằng tiếng nước ngoài. Chúng tôi đã cố gắng giải thích thêm thuật ngữ trong phần phụ lục. Có những thuật ngữ tiếng Việt mà chúng tôi chưa biết chính xác, có thể dịch sai, rất mong bạn đọc thông cảm và cho biết ý kiến.

Mục đích của chúng tôi xuất bản quyển sách này là trao đổi kiến thức cơ bản trong lĩnh vực sinh học phân tử, ứng dụng vào di truyền học và chọn giống, một cách nhanh chóng và cập nhật hóa. Bởi vì sự phát triển của công nghệ sinh học rất đa dạng, khối lượng thông tin ngày một lớn, cần được hệ thống rõ ràng.

Nhiều năm trước đây, kỹ thuật lập bản đồ gen cây trồng là một công cụ mới có hiệu quả trong nghiên cứu sinh học thực vật, cũng như nhiều môn học khác. Nó ra đời sau hàng loạt các tiến bộ kỹ thuật DNA căn bản khác như kỹ thuật cloning. Từ đó nó giúp cho người ta giảm đi rất nhiều các bước nghiên cứu thường lệ, và nhiều kỹ thuật mới trong cloning ở mức độ phân tử ra đời, như một kết quả trực tiếp của nhu cầu "phân tích genome". Một vài khái niệm mới cũng đi kèm theo như "map-based cloning" đang được ứng dụng trong chọn giống nhờ marker phân tử. Trong nhiều năm qua, người ta tập trung mọi nỗ lực cho phân

tích genome của người, thông qua dự án nổi tiếng có tên là "Human genome project". Hiện nay, yêu cầu phát triển của nông nghiệp trước nguy cơ cạn kiệt nguồn tài nguyên thiên nhiên và sự phát triển dân số khó kiểm chế, người ta có xu hướng đầu tư nhiều cho phân tích genome thực vật, làm nền tảng cho một cuộc cách mạng xanh mới, giải quyết bài toán lương thực cho thế giới. Phân tích genome thực vật hay genome cây trồng chắc chắn sẽ đóng góp những thay đổi quan trọng trong cải tiến giống, và làm phong phú thêm kiến thức của loài người về di truyền phân tử. Hi vọng quyển sách này sẽ đáp ứng phần nào cho mục tiêu mà chúng ta đề cập.

BÙI CHÍ BỬU - NGUYỄN THỊ LANG

MỤC LỤC

Trang

Mở đầu	5
Chương I: Cấu trúc và chức năng phân tử DNA	9
1-1. Nhân và nhiễm thể	11
1-2. Tự tái bản DNA	14
1-3. Gen là đơn vị có tính chất đột biến	17
1-4. Số lượng DNA trong nhân và giá trị C	18
1-5. Cấu trúc của chuỗi mã DNA	19
1-6. DNA vệ tinh	27
1-7. Chức năng của chuỗi mã DNA có tính lặp lại	27
1-8. Gen mã hóa trong tổng hợp protein	29
1-9. Các nguyên tố chuyển vị	31
Chương II: Hoạt động của các enzyme trong phân tích genome	35
2-1. Thông tin di truyền	35
2-2. Cấu trúc enzyme	37
2-3. Ảnh hưởng của cấu trúc đến chức năng enzyme	41
2-4. Restriction endonuclease	45
Chương III: Những vectơ thông dụng trong kỹ thuật di truyền	63
3-1. Plasmid	63
3-2. Vectơ Lambda của thực khuẩn thể	70
3-3. Cosmid	74
3-4. YAC	77
3-5. BAC	80
Chương IV: Liên kết giữa gen và marker	93
4-1. Các loại DNA marker	93
4-2. RFLP marker	95
4-3. Marker là sản phẩm của PCR	103
4-3-1. Nguyên tắc cơ bản của PCR	103
4-3-2. RAPD marker	107
4-3-3. SSCP marker	109
4-3-4. STS marker	110
4-3-5. Microsatellite marker	112
4-3-6. AFLP marker	116
Chương V: Phương pháp tính giá trị liên kết gen	123
5-1. Khoảng cách giữa hai gen	123
5-2. Dự đoán recombination	124
5-2-1. Phương pháp hồi giao	124
5-2-2. Phương pháp phân tích quần thể F_2	126

5-2-3. LOD scores	128
5-3. Thực hành	128
Chương VI: Kỹ thuật nhân bản vô tính (cloning)	137
6-1. Phương pháp cloning	137
6-2. Kho lưu trữ DNA của genome (library)	137
6-3. Phân tích chuỗi mã DNA có phân tử lớn và chromosome walking với khoảng cách rộng	144
6-4. Kho lưu trữ cDNA	147
Chương VII: Bản đồ di truyền & bản đồ vật lý	159
7-1. Khái niệm bản đồ trong di truyền cổ điển	159
7-2. Bản đồ liên kết gen và marker phân tử	159
7-3. Bản đồ vật lý	168
7-4. Áp dụng trên cây lúa (case study)	174
Chương VIII: Phản ứng chuỗi polymersae (PCR)	195
8-1. Nguyên tắc chung	195
8-2. Tối ưu hóa các điều kiện cho phản ứng PCR	199
8-3. Định lượng DNA và RNA bằng PCR	203
8-4. Thiết kế primer	205
8-5. Những ứng dụng quan trọng của PCR	209
8-5-1. Tạo nhanh các gen tổng hợp	209
8-5-2. Kỹ thuật cloning cDNA bằng PCR đảo	211
Chương IX: Chuỗi mã di truyền DNA (DNA sequence)	217
9-1. Nguyên tắc cơ bản của DNA sequencing	217
9-2. Chuẩn bị dây template	223
9-3. Điện di trên gel	224
9-4. Phương pháp không dùng chất phóng xạ	225
9-5. Phân tích computer chuỗi mã DNA và protein	228
9-6. Kỹ thuật sequencing trực tiếp đối với sản phẩm PCR	231
Chương X: Chọn lọc giống cây trồng nhờ marker phân tử DNA	239
10-1. Hồi giao	240
10-2. Bản đồ di truyền	241
10-3. Phân tích QTL	244
10-4. Bản đồ vật lý	245
10-5. Áp dụng bản đồ di truyền trong chọn tạo giống	246
10-6. Quan hệ giữa chọn giống cổ điển và chọn giống nhờ marker	247
10-6-1. Liên kết gen và số loci của marker	247
10-6-2. Chọn lọc cá thể trong quần thể tạp giao	248
10-6-3. Phương sai di truyền về tính cộng được giải thích bằng marker	250
10-6-4. Chọn lọc gia phả	251
10-6-5. Tuyển chọn tái tục	253
10-6-6. Thảo luận mối quan hệ giữa chọn giống cổ điển và chọn giống nhờ marker	255
PHỤ LỤC: Thuật ngữ chuyên môn	261

Chương I

CẤU TRÚC VÀ CHỨC NĂNG PHÂN TỬ DNA

Có ba dạng chức năng trong thông tin di truyền :

- Thông tin được liên tục nhờ sự tự tái bản [**replication**]; một nucleic acid nào đó trong dây đôi được sao lại [**duplicated**] để có những copies giống nhau.
- Thông tin được thể hiện bởi tiến trình có tính chất hai giai đoạn
- Sự chuyển mã [**transcription**] tạo ra một dây đơn RNA có cùng sequence với một trong những dây đơn của duplex DNA.
- Sự giải mã [**translation**] làm cho nucleotide sequence của RNA này trở thành sequence của amino acids - thành viên của protein nào đó.

Thông tin di truyền được lưu trữ dưới dạng các nucleotide triplet hay còn gọi là codon. Mỗi codon bao gồm ba nucleic acid, mang một amino acid tương ứng.

Chúng ta có thể chia dây đôi DNA ra thành hai phần:

- Dây đơn mang cùng sequence với mRNA được gọi là coding strand hoặc sense strand.
- Dây đơn đối xứng được gọi là template strand hoặc antisense strand, nơi đó xảy ra sự tổng hợp trực tiếp của mRNA thông qua sự bắt cặp bổ sung các base.

Chúng ta còn dùng thuật ngữ antisense để diễn tả một sequence của DNA hoặc RNA trở thành bổ sung đối với mRNA.

Vì mã di truyền được đọc trên mRNA, cho nên bốn base thường hiện diện sẽ là U, C, A và G

RNA mang [tRNA] là adaptor

Mỗi amino acid được đại diện bởi ít nhất một codon, bao gồm một nucleotide triplet trong mRNA. Việc giải mã có thể được thực hiện bởi một phân tử có tính chất adaptor. Adaptor này chính là tRNA, một phân tử nhỏ mà chuỗi polynucleotide chỉ có khoảng 75-85 bases.

tRNA có hai đặc điểm:

- Nó chỉ thể hiện một amino acid được liên kết đồng hóa trị
- Nó có chứa một nucleotide sequence- còn gọi là anticodon - bổ sung cho codon

của amino acid. Thẻ anticodon làm cho tRNA có thể ghi nhận codon thông qua sự bắt cặp bổ sung các base

tRNA có cấu trúc cấp hai, với 4 nhánh. Nucleotide sequence của mỗi tRNA có thể được viết theo dạng cloverleaf, trong đó, mỗi base bổ sung tạo thành các nhánh thân [stems], mỗi nhánh được cấu tạo bởi:

- Dây đơn hình vòng (single strand loop)
- Thân mang các cặp base (base-paired stem)

Khi một tRNA mang theo nó một amino acid tương ứng với anticodon, nó sẽ trở thành **aminoacyl-tRNA**. Amino acid này được gắn vào nhờ cầu nối ester tại gốc carboxyl với gốc hydroxyl của ribose- vị trí 2' và 3'

Có ít nhất một tRNA (nhưng thường nhiều hơn một) cho mỗi amino acid.

Một tRNA có tên kèm theo 3 chữ viết tắt của amino acid. Thí dụ tRNA^{Ala} là tRNA đối với alanine. Nếu có nhiều hơn một tRNA cho cùng một amino acid, người ta đánh số để phân loại chúng. Thí dụ tRNA₁^{Tyr}, tRNA₂^{Tyr} là 2 tRNA đối với tyrosine.

Một tRNA mang một amino acid còn được ký hiệu bởi một tiếp đầu ngữ mang tên amino acid ở phía trước. Thí dụ Ala-tRNA.

Quá trình mang của tRNA được xúc tác bởi enzyme đặc biệt aminoacyl - tRNA synthase. Có ít nhất 20 aminoacyl - tRNA synthase. Mỗi enzyme chuyên biệt đối với một amino acid.

RNA thông tin [mRNA] được giải mã nhờ ribosome:

Protein được tổng hợp từ C-terminus đến N-terminus. Việc tổng hợp amino acid thành protein được thực hiện tại ribosome. Một ribonucleoprotein bao gồm hai subunits. Mỗi subunit của ribosome gồm có nhiều protein kết hợp với phân tử rất dài RNA. RNA này được gọi là ribosome RNA, viết tắt là rRNA.

Ribosome được mô tả theo nghĩa mức độ lắng đọng [sedimentation] của nó, đo bằng chỉ số Svedbergs, trong đó nếu giá trị S càng cao thì mức độ lắng đọng càng lớn và khối lượng cũng lớn. Ribosome của vi khuẩn có sediment gần bằng 70S, ribosome của thực vật bậc cao có sediment khoảng 80S.

Một ribosome gồm có hai subunits. Hai subunits tách rời ra khi hàm lượng ion Mg⁺⁺ giảm. Ngược lại, nếu thêm vào Mg, hai mảnh này sẽ gắn với nhau trở lại. Ribosome của vi khuẩn (70S) có những subunits với sediment 50S và 30S. Còn subunits của tế bào chất thực vật bậc cao 80S có sediment là 60S và 40S. Hai subunits này cùng hoạt động như là một phần của ribosome hoàn chỉnh, nhưng có nhiệm vụ khác nhau trong sinh tổng hợp protein. Tất cả các ribosome có trong một tế bào đều giống nhau. Chúng có nhiệm vụ tổng hợp các

protein khác nhau bằng cách phối hợp với những mRNA khác nhau, những mRNA này cung cấp các mật mã di truyền.

Ribosome tạo ra môi trường kiểm soát sự ghi nhận giữa một codon của mRNA và anticodon của tRNA. Để hoàn thiện sinh tổng hợp protein, ribosome phải di động theo mRNA, một codon trong một thời gian nhất định.

Một ribosome dính với mRNA tại vị trí 5' của một vùng mang mật mã; rồi di chuyển theo RNA này cho đến vị trí 3', nó giải mã trên mỗi triplet codon thành một amino acid trên đường đi. Trong tiến trình này, aminoacyl-tRNA tương ứng sẽ phối hợp với nó, điều khiển amino acid tổng hợp thành chuỗi polypeptide. Tại một thời điểm nào đó, ribosome có thể định vị hai aminoacyl-tRNA tương ứng với những codon, tạo điều kiện dễ dàng cho việc thành lập cầu nối peptide giữa hai amino acid tương ứng. Dần dần, chuỗi polypeptide sẽ trở nên dài hơn một amino acid.

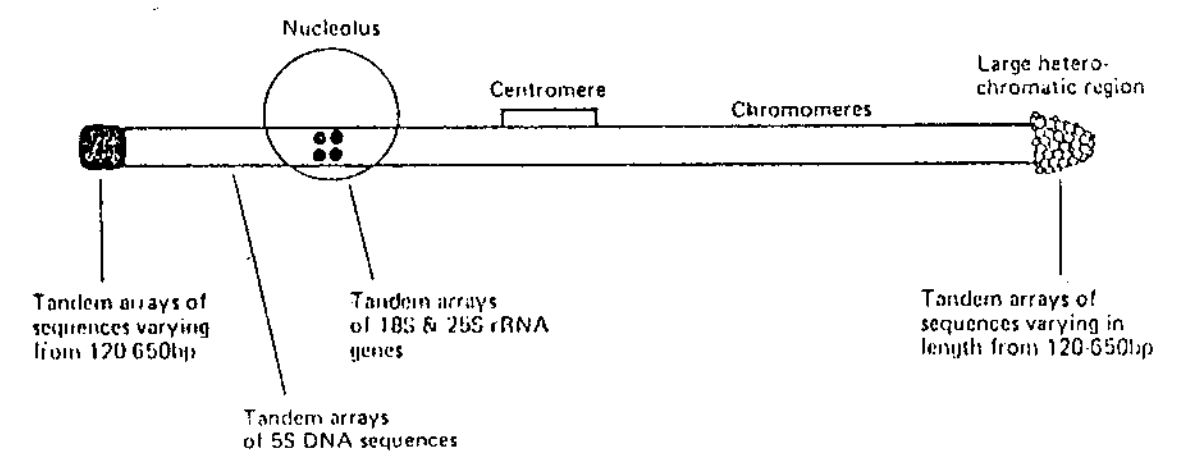
Khi những ribosome hoạt động được xác định, dưới dạng kết hợp với protein mới được tổng hợp, người ta sẽ tìm thấy dạng của một unit bao gồm một mRNA kết hợp với nhiều ribosome. Người ta gọi đó là polyribosome hay polysome. Mỗi ribosome trong polysome này tổng hợp một cách độc lập một polypeptide đơn trong khi lướt qua chuỗi mã có tính chất thông tin. Như vậy, một mRNA có một series các ribosome, nó mang protein mới tạo ra, di chuyển từ vị trí 5' đến 3', mà chiều dài của protein này ngày một dài thêm. Chuỗi polypeptide trong quá trình sinh tổng hợp như vậy còn được gọi là **nascent protein** [protein mới sinh].

1.1. NHÂN VÀ NHIỄM THỂ

Nhân của thực vật bậc cao có thể có một nồng độ khoảng 100 mg.ml^{-1} phân tử DNA. Đây chính là hoạt động chủ yếu của thông tin di truyền, điều khiển các hoạt động sinh trưởng, phát triển của sinh vật. Việc tìm hiểu cấu trúc, chức năng của DNA, cũng như làm thế nào nó thể hiện việc điều tiết các chức năng, là quá trình nghiên cứu liên tục của ngành học di truyền phân tử

Nhân của thực vật bậc cao thường có kích thước 3-20 μm , trường hợp của tảo *Acetabularia* có thể đạt kích thước 150 μm . Nhân được bao bọc bởi hai lớp màng nhân, có những lỗ rất nhỏ, đường kính 50-100 nm. Những lỗ này rất năng động, chứa một hỗn hợp protein xúc tác làm cho những đại phân tử có thể qua lại màng nhân. Số lượng và vị trí của những lỗ này thay đổi tùy theo quá trình tăng trưởng và phân hóa một cách nhanh chóng. Điều này phản ánh bản chất năng động của màng nhân. Nó nối liền với mạng võng nội chất (endoplasmic reticulum), đôi khi người ta còn có thể thấy sự phối hợp chặt chẽ giữa màng nhân và màng của ti thể bộ hoặc lạp thể. Trong khi nhân phân cắt, màng nhân tạm thời biến mất vào lúc nhiễm thể tách đôi. Hạch nhân (nucleolus) xuất hiện sau khi gián phân tại những vị trí đặc biệt trên nhiễm thể, còn được gọi là nucleolus organizers (hình 1-1). Nó

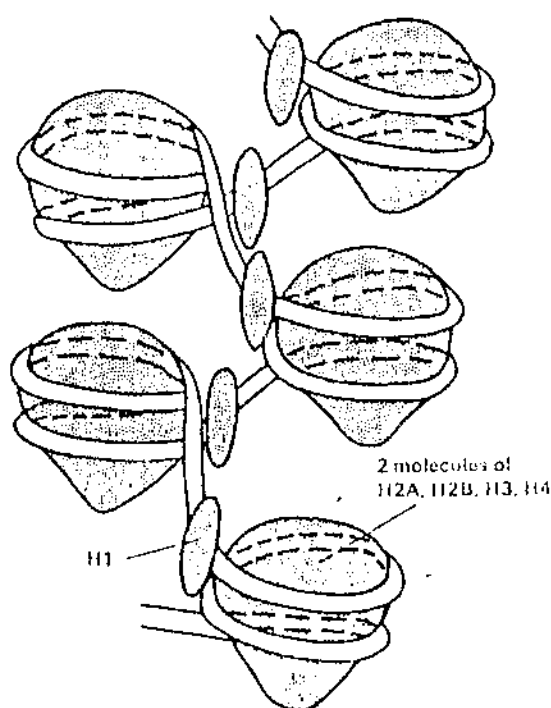
phát triển thành những cơ quan rõ nét, không có màng giới hạn nào cả, chứa các DNA và sợi + hạt RNA và protein. Đó là vị trí của giải mã của rRNA.



Hình 1.1 : Vị trí của chuỗi mã DNA được lập lại trên nhiễm thể IR của kiểu mạch, quan sát trong giai đoạn pachytene của tiền kỳ (nguyên phân)

Nhân chứa các protein cấu trúc (tubulin, actin), polymerase enzyme, protein có tính acid làm nhiệm vụ điều tiết, RNA và những protein có tính kiềm như histone. Tỷ lệ DNA : histone : RNA : acidic protein là 1 : 1 : 0,1 : 0,6 để tạo ra chromatin (chất nhiễm sắc). **Đơn vị có tính lặp lại của chromatin được gọi là nucleosome.** Nó chứa một chùm đặc biệt các histone kết hợp với DNA. Có năm histone protein là: H1, H2A, H2B, H3 và H4. Histone H1 rất giàu lysine, H2A và H2B nghèo lysine hơn, H3 và H4 rất giàu arginine. Cấu trúc sơ cấp của H3 và H4 rất giống nhau trong nhiều sinh vật, thí dụ như chuỗi mã amino acid của H4 trong đậu trắng (pea) và đậu Hà Lan (cow pea) gần như là một. Trái lại, H2A và H2B dường như biến thiên nhiều hơn. Tương tự, H1 được nhiều nghiên cứu cho thấy có sự biến thiên rất cao.

Đặc điểm chung của cấu trúc nucleosome là thể hiện giống nhau trong mọi sinh vật có nhân thật eukaryotes (sinh vật bậc cao). Những histone được kết chùm do phản ứng được gọi là hydrophic interaction của vùng mang carbon cuối cùng của nó. Điều này còn do một loạt các phản ứng xen kẽ của phân tử mang điện âm và dương nhằm ổn định cấu trúc, hình thành hiện tượng được gọi là "salt bridges". Đoạn base chứa N ở vị trí cuối trong protein được thể hiện ở bề mặt của phân tử. Nó phản ứng với gốc phosphate của DNA xoắn (helix) tạo ra một phân tử DNA xoay về bên trái 1,8 lần xung quanh mặt ngoài của phân tử. Histone H1 không có trong nucleosome nhưng phản ứng với linker DNA (hình 1.2). Chiều dài của linker DNA dường như hơi thay đổi giữa những sinh vật khác nhau, những mô khác nhau.



Hình 1.2 : Hiện tượng quấn theo vòng của DNA xung quanh histone protein để tạo ra nucleosome

Cấu trúc của nucleosome được nghiên cứu nhờ sự trợ giúp của các phương tiện hóa học, kính hiển vi điện tử, X-quang, và sự miễn dịch của DNA đối với phản ứng phân cắt nhờ enzyme (nuclease). Chính phản ứng với các nuclease cho phép chúng ta nghiên cứu sự sắp xếp của DNA.

Kết quả chính của sự phối hợp DNA với nucleosome làm giảm chiều dài tổng thể của nó còn 1 / 7 so với phân tử gốc, tạo ra một sợi có độ dày khoảng 10nm. Sự nén lại như vậy ít hơn rất nhiều lần trong nhiễm thể ở giai đoạn metaphase (trung kỳ). Điều này có nghĩa là: một cấu trúc có mức độ sắp xếp cực kỳ tinh vi, cho phép chứa đựng phân tử DNA có chiều dài hơn 1m, nằm gọn trong một nhân có kích thước bé hơn rất nhiều lần. Giai đoạn đầu của quá trình nén bao gồm tiến trình tạo xoắn của nucleosome ở dạng solenoid có đường kính 30nm, cho phép một tiến trình nén tiếp theo. Phản ứng phosphoryl hóa của H1 được xem như đóng góp vai trò quan trọng trong tiến trình nén chromatin, khi tế bào gián phân. Việc cải tiến của các histone khác có thể tạo ra ảnh hưởng đến cấu trúc chromatin, nhưng chúng ta chưa có đủ bằng chứng xem nó xảy ra như thế nào. Chúng ta cũng chưa biết rõ cơ sở sinh học phân tử để phân biệt theo mức độ tế bào giữa hai vùng nhiễm sắc khác nhau của chất nhiễm sắc thật (euchromatin) và chất dị nhiễm sắc (heterochromatin), mặc dù người ta đã biết rằng chất dị nhiễm sắc có độ nén cao hơn, và

chứa đựng đoạn mã lặp lại ở tần suất cao, nhưng đoạn mã này không chuyển được trong RNA (hình 1.1). Nếu tương tác giữa DNA và histone có tính chất tĩnh điện (electrostatic), thì sự hình thành và định vị của nucleosome sẽ có tính chất độc lập đối với chuỗi mã DNA. Tuy nhiên, có những chứng cứ cho thấy trên động vật, những DNA vệ tinh (satellite DNAs) có những chuỗi mã lặp lại từng đoạn rất đều đặn, chúng phân bố có tính chất không ngẫu nhiên trong nucleosome. Điều này chứng tỏ rằng chuỗi mã này có thể làm ảnh hưởng đến sự định vị hoặc sự sắp xếp nucleosome.

Người ta chưa có nguyên tắc chung về sự sắp xếp DNA trong nucleosome, mặc dù bằng các polymerase enzyme, người ta đã sao chụp được chúng. Có thể trong khi chuyển mã (transcription) hoặc tự tái bản (replication), những nucleosome này phân rã hoặc chuyển đổi vị trí trên dây đơn DNA. Người ta thấy rằng, DNA vừa được tổng hợp rất miễn cưỡng với nuclease và có thể không có hình dáng một nucleosome bình thường trong quá trình tự tái bản.

1.2. TỰ TÁI BẢN DNA

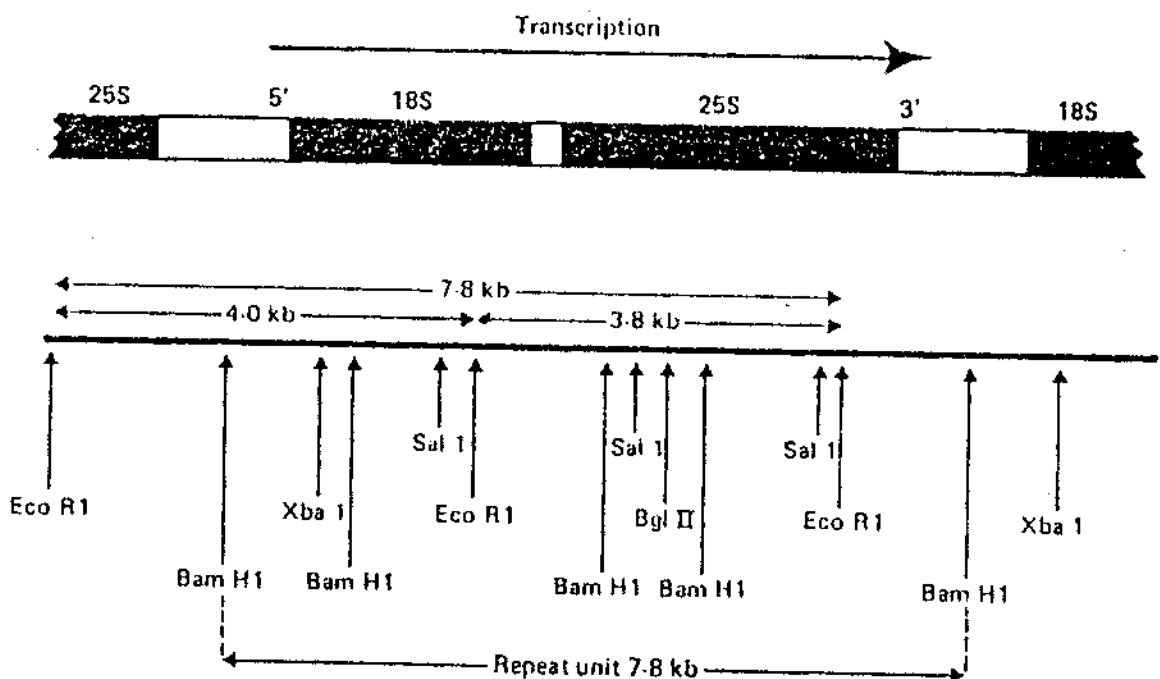
Hầu hết các tế bào thực vật hoàn tất chu kỳ sinh trưởng và gián phân trong vòng 15-40 giờ, tùy theo loài thực vật và nhiệt độ môi trường. Sự tự tái bản DNA và sự tổng hợp histone khoảng 7-11 giờ, hiện tượng này góp phần trong kỳ nghỉ phân bào (interphase), còn được gọi là S phase.

Sinh tổng hợp DNA đã được nghiên cứu trên cây được đánh dấu bằng phóng xạ. Kết quả cho thấy mỗi phân tử được tái bản hàng trăm lần. Tiến trình tự tái bản theo cả hai hướng, xuất phát từ điểm gốc của tái bản (origin). Mỗi nhánh DNA được tổng hợp gọi là "replicon". Các nhà nghiên cứu đã tìm thấy những replicon của thực vật bậc cao thường có chiều dài từ 20 đến 30 μ m hoặc 60-90 kilo base-pairs(kp). Có 2-25 họ của replicon trong thực vật bậc cao, mỗi replicon trong một họ trải qua tiến trình sinh tổng hợp DNA cùng một thời điểm, trong S phase, nhưng nếu khác họ thì nó xảy ra ở các thời điểm khác nhau.

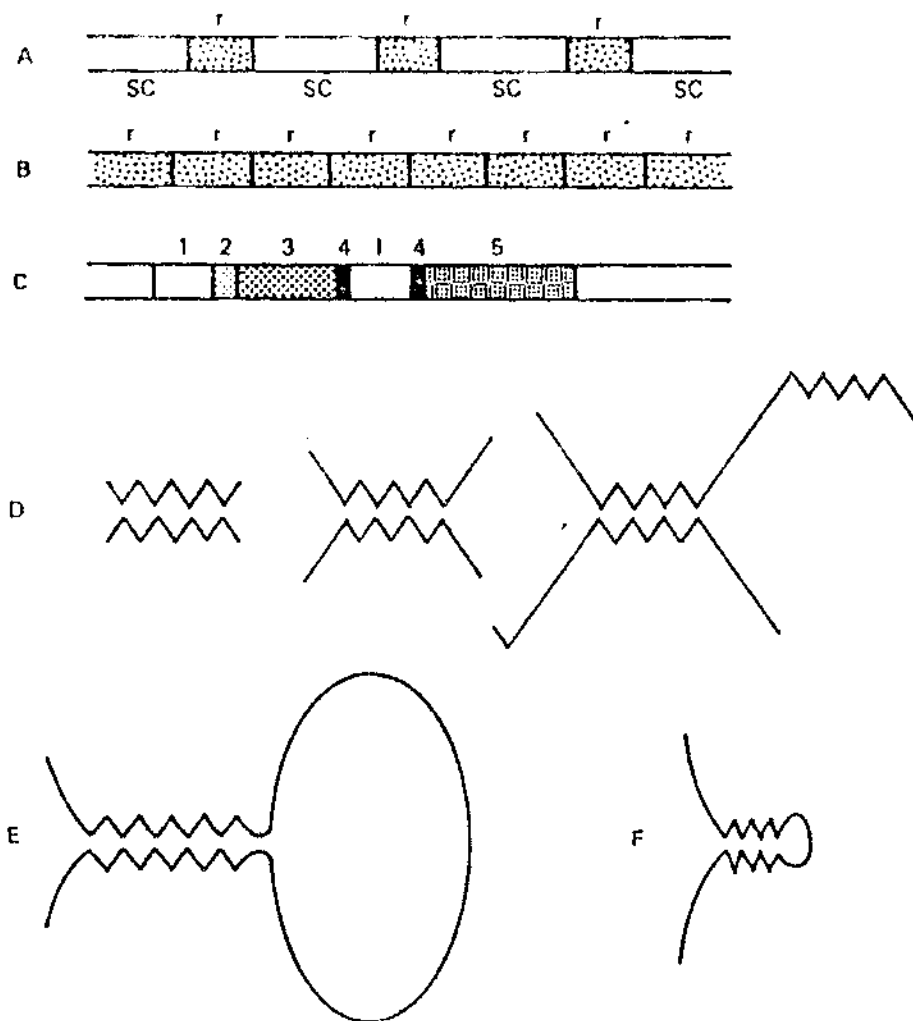
Sự tự tái bản DNA là một tiến trình có tính chất bảo thủ phân nửa (semi-conservative), với các dây đơn DNA bổ sung được tổng hợp theo hướng 5' - 3' do hoạt động của DNA polymerase, sử dụng các dây đơn khác như dây template (dây mang bản sao ngược với bản gốc). Người ta không thể tổng hợp cả hai dây mới này liên tục bởi vì enzyme polymerase chỉ hoạt động theo hướng 5' - 3'. Sự tổng hợp không liên tục này của cả hai dây xảy ra trong vài trường hợp, nhưng thông thường người ta tin rằng sự tự tái bản DNA bao gồm việc tự tái bản có tính không liên tục một nửa (semi-discontinuous), với một dây gốc (leading strand) được tổng hợp theo hướng 5' - 3', và một dây đối xứng kéo dài theo hướng 3' - 5' do sự gắn lại của những đoạn ngắn DNA.

Sự tự tái bản DNA trong thực vật bao gồm hoạt động của men α - và γ -DNA polymerase dạng dung dịch hòa tan. Men β -polymerase gắn với chromatin có thể có trong

khi DNA bắt cặp đôi trở lại. Sự hiểu biết của chúng ta về sinh hóa học của việc tự tái bản DNA ở thực vật bậc cao vẫn còn rất ít. Hình 1.3 cho thấy nghiên cứu tự tái bản ở động vật và ở sinh vật eukaryote bậc thấp (Bryant 1982). (1) Tổng hợp DNA được bắt đầu ở điểm gốc (replication fork) bằng cách sinh sản và xoắn ngược DNA nhờ men topoisomerase còn được gọi là "unwinding enzyme" hay "*helicase*". (2) Những dây đơn DNA khác nhau được hình thành nhờ men có tên là "single-stranded DNA-binding enzyme". (3) Một đoạn mã RNA ngắn được tổng hợp tại vị trí gốc (origin), kế đến, nó được sử dụng làm primer cho các sinh tổng hợp tiếp theo sau, trên dây gốc (leading strand) theo hướng 5' - 3'. (4) Trên dây đối nghịch lại, hoạt động của men *helicase* được tiếp nối bởi sự phát sinh một "*primase*" của đoạn ngắn RNA (primer), dùng trong tiến trình tổng hợp không liên tục. (5) DNA polymerase làm kéo dài RNA primer của dây đối nghịch theo hướng 5' - 3' để sản xuất ra DNA mới (nascent), còn được gọi là "Okazaki fragment" (tên của người đã tìm ra hiện tượng này đầu tiên). Những đoạn này có khoảng 200 nucleotide, đã được tìm thấy trên đậu nành và *Physarum sp.* (6) RNA primers được lấy ra từ Okazaki fragment nhờ men *nuclease*. (7) Những khoảng trống trên DNA còn lại (gaps) được bắt cặp trở lại bởi men *DNA polymerase* thứ cấp. (8) Những đoạn DNA kế bên được nối dính lại với nhau thành ra một dây mới gọi là "new daughter strand". (9) Những RNA primers mới và những đoạn Okazaki được tổng hợp thành những điểm gốc tự tái bản (fork) tiếp theo. (10) Các dây DNA vừa mới được tổng hợp từ các replicon kế cận được nối liền với nhau.



Hình 1.3 : Vị trí của các enzym phân cắt hạn chế trong rRNA của đậu nành. rRNA này có trong nhân, có tính chất lặp lại định kỳ (tandemly-repeated). Chuỗi mã 5.8S rRNA nằm giữa 18S và 25S rRNA



Hình 1.4 : Hình thái chuỗi và đặc tính tái tổ hợp của nhiều loại hình repeated DNA. (A) Sự phân bố rải rác của một sequence giữa DNA có bản sao đơn (single-copy). (B) Chuỗi mã DNA lặp lại có định kỳ. (C) DNA Lặp lại với nhiều loại hình khác nhau. (D) Phát hiện DNA lặp lại giữa những "single-copy DNA". (E) Hiện tượng hoàn nguyên từ bên trong của chuỗi mã ngược khác với single-copy. (F) Giống như (E) nhưng là chuỗi mã gắn ké bên, đường zig-zag cho biết vùng base bất cặp nhau.

Người ta có khoảng 200 nucleotide Okazaki fragment, nó có xu hướng ngắn hơn trong sinh vật prokaryotype (như vi khuẩn), nó góp phần trong việc lý giải sự sắp xếp DNA trong nucleosome. Nghiên cứu về sự tự tái bản trong virus như adenovirus của tế bào động vật cho thấy: xuất phát điểm của sinh tổng hợp DNA từ những chuỗi mã có tính lặp lại ngược (inverted repeat) là TTGGATTGAAGCCAA. Những chuỗi mã như vậy có tính đối xứng tuần hoàn, và tạo thành DNA hình vòng (hairpin loop), nó có thể được dùng làm tín hiệu cho sự bắt đầu tự tái bản.

Sau khi tự tái bản ở thực vật bậc cao, hơn 25% cytosine được methyl hóa do *DNA methylase* mà nó sử dụng S-adenosyl methionine làm chất cho methyl. Trong vài trường hợp, phản ứng methyl hóa ở một vị trí chuyên biệt nào đó trở nên vô cùng quan trọng trong việc điều tiết sự thể hiện của gen.

1.3. GEN LÀ ĐƠN VỊ CÓ TÍNH CHẤT ĐỘT BIẾN

Kiểu gen bao gồm một bộ hoàn chỉnh những thông tin di truyền được sinh vật truyền đi từ thế hệ này sang thế hệ khác. Biểu hiện ra bên ngoài của kiểu gen là kiểu hình, đó là trạng thái vật chất của sinh vật. Kiểu gen bao gồm nhiều gen, có trong nhiễm thể. Theo định luật Mendel, gen được xem xét dưới nhiều yếu tố tiềm ẩn như : các alen phân ly và các gen phân ly một cách độc lập. Tiến trình gián phân giảm nhiễm cho thấy cơ sở có tính vật chất của trạng thái theo các định luật Mendel.

Gen trên các nhiễm thể khác nhau, hoặc gen cách xa nhau tuy ở trên cùng một nhiễm thể, thì tái tổ hợp một cách độc lập. Những gen trên cùng một nhiễm thể tạo nên một nhóm liên kết gen theo kiểu đường thẳng (linear), trong đó, gen gần nhau có xu hướng tạo ảnh hưởng di truyền chung với nhau. Bản đồ liên kết gen cho thấy sự biểu hiện theo đường thẳng đối với các vị trí mà gen định vị trên nhiễm thể. Vật liệu di truyền của một nhiễm thể tạo nên một cấu trúc có tính chất liên tục, trong đó, gen có cùng một kiến trúc theo đường thẳng.

Wild type mô tả kiểu gen hoặc kiểu hình có ích. Một gen có tính chất wild type sản sinh ra một protein có chức năng [functional]. Các đột biến là những biến đổi có tính di truyền trong thông tin di truyền. Gen đột biến có thể tạo ra một protein bị biến đổi, hoặc không hoàn thành việc tạo ra một protein có chức năng. Một thể null mutant sẽ không tạo ra được một protein nào cả. Nếu chức năng của một alen đơn nào đó đã đầy đủ trong một tế bào $2n$, thì alen có tính wild type sẽ trội so với mutant có tính lặn. Tính trội không hoàn toàn [partial dominance] sẽ xảy ra khi các chức năng của gen có tính số lượng [hai alen sẽ tạo ra ảnh hưởng gấp đôi so với một alen]. Đồng tính trội [codominance] là kết quả của hai alen khác nhau, có sự chuyên biệt khác nhau, sao cho thể dị hợp xuất hiện được cả hai đặc tính của bố và mẹ. Đột biến có thể trở nên có tính chất điều kiện [conditional], khi thể hiện kiểu hình mutant dưới những điều kiện không cho phép, nhưng lại thể hiện "wild type" dưới những điều kiện được phép.

Mối quan hệ giữa hai đột biến gần nhau trên bản đồ, có cùng một kiểu hình, có thể được xác định bằng phép thử đồng-đối [cis-trans test]. Nếu đột biến không hoàn thành việc bổ sung trong thể trans configuration, nó sẽ được thể hiện ở thể cistron, hoặc gen. Một cistron cũng tạo ra mã [codes] cho một chuỗi polypeptide đơn nào đó. Để khỏi làm cho khái niệm gen trở nên phức tạp, khi nói rằng một gen có thể bao gồm nhiều gen nhỏ hơn, người ta sử dụng thuật ngữ cistron. Như thế một gen có thể gồm nhiều cistron, mỗi cistron có chức năng khác nhau.

1.4. SỐ LƯỢNG DNA TRONG NHÂN VÀ GIÁ TRỊ C

Bảng 1.1 : Giá trị 1C của DNA trong nhân của thực vật bậc cao so với virus, chloroplast, bacterium và người

	DNA gram/genome	Độ bội thể	Số cặp base (bp) trong genome	Trọng lượng phân tử
Virus khảm trên cải bông	$0,84 \times 10^{-17}$		8042	$5,1 \times 10^6$
Pea chloroplast	$1,3 \times 10^{-16}$		$1,24 \times 10^5$	80×10^6
<i>Escherichia coli</i>	$4,0 \times 10^{-15}$		$3,9 \times 10^6$	$2,5 \times 10^9$
<i>Allium cepa</i> (hành tây)	$16,8 \times 10^{-12}$	$2x = 16$		
<i>Arabidopsis thaliana</i>	$0,2 \times 10^{-12}$	$2x = 10$		
<i>Avena sativa</i> (oat)	$13,7 \times 10^{-12}$	$6x = 42$		
<i>Brassica napus</i> (turnip)	$1,6 \times 10^{-12}$	$2x = 38$		
<i>Beta vulgaris</i> (beet)	$1,2 \times 10^{-12}$	$2x = 18$		
<i>Curcubita melo</i> (melon)	$1,0 \times 10^{-12}$	$2x = 24?$		
<i>Glycine max</i> (đậu nành)	$0,9 \times 10^{-12}$	$4x = 40$		
<i>Gossypium hirsutum</i>	$3,0 \times 10^{-12}$	$4x = 52$		
<i>Hordeum vulgare</i>	$5,6 \times 10^{-12}$	$2x = 14$		
<i>Lilium davidii</i>	$43,2 \times 10^{-12}$	$2x = 24$		
<i>Lolium perenne</i>	$4,9 \times 10^{-12}$	$2x = 14$		
Tomato	$0,75 \times 10^{-12}$	$2x = 24$		
<i>Nicotiana sylvestris</i>	$2,2 \times 10^{-12}$	$2x = 24$		
<i>Nicotiana tabacum</i>	$3,9 \times 10^{-12}$	$4x = 48$		
<i>Oryza sativa</i>	$1,0 \times 10^{-12}$	$2x = 24$		
<i>Pisum sativum</i> (pea)	$4,9 \times 10^{-12}$	$2x = 14$		
<i>Raphanus sativus</i>	$0,4 \times 10^{-12}$	$2x = 18$		
<i>Secale cereale</i>	$9,5 \times 10^{-12}$	$2x = 14$		
<i>Solanum tuberosum</i>	$2,1 \times 10^{-12}$	$4x = 48$		
<i>Spinacia oleracea</i>	$1,0 \times 10^{-12}$	$2x = 12$		
<i>Triticum monococcum</i>	$6,2 \times 10^{-12}$	$2x = 14$		
<i>Triticum aestivum</i>	$17,3 \times 10^{-12}$	$6x = 42$		
<i>Vicia faba</i>	$13,3 \times 10^{-12}$	$2x = 12$		
<i>Zea mays</i>	$3,9 \times 10^{-12}$	$2x = 20$		
Người	$6,0 \times 10^{-12}$	$2x = 46$		

1 picogram (10^{-12} g) = $0,965 \times 10^9$ bp, hoặc $6,4 \times 10^{11}$ daltons

Số lượng DNA trong nhân của thực vật được ước đoán nhờ kỹ thuật phân tích hóa học trực tiếp, tùy theo số nhân được lấy mẫu, theo sau đó là đo tỉ trọng ở dạng micro các nhân được nhuộm phẩm Feulgen. Những giá trị này thường thể hiện ở dạng 1C hay 2C của phân tử DNA, trong đó 1C là hàm lượng DNA chưa được tái bản có tính chất đơn bội (unreplicated haploid) (xem bảng 1.1). Hàm lượng DNA trong nhân của các loài thực vật bậc cao thay đổi từ 0,5 đến 200 picograms (Bennett và Smith 1976, Bennett và ctv. 1982). Đậu trắng (pea) chứa DNA cao gấp 1000 lần hơn *E.coli*, và *E.coli* có thể chứa khoảng 4000 gen. Nếu tất cả DNA của đậu pea được thể hiện theo kiểu cổ điển, nó sẽ có bốn triệu gen. Người ta rất khó ước đoán số gen cần thiết đóng góp vào sự khác biệt về mặt sinh học giữa pea và *E. coli*, như vậy có một giá trị cho thấy thông thường thực vật có chuỗi mã di truyền gấp 20 lần so với người, ước khoảng 6×10^{-12} g DNA, và phức tạp gấp nhiều lần hơn. Sinh vật nào có số gen cần thiết lớn, thì nó sẽ có xác suất bị đột biến cao hơn, làm ảnh hưởng đến kiểu hình của nó.

Hàm lượng DNA đơn bội ở nhân của thực vật bậc cao cũng được ước đoán bằng cách nghiên cứu động thái của sự tái tạo lại DNA (renaturation). Đối với cây đậu pea, thuốc lá, đậu nành, kết quả cho thấy có sự giống nhau về phương diện hóa học và tế bào học. Kết quả phân tích động thái còn cho thấy một tỉ lớn khá lớn DNA có trong những chuỗi mã có tính tự tái bản cao, khác với những gen đặc trưng cho sự mã hóa protein. Tỉ lệ DNA của thực vật bậc cao xuất hiện trong một số copy (có tên gọi là "single-copy DNA") thay đổi từ 20-80%. Xu hướng chung ở thực vật là genome càng lớn, DNA tự tái bản càng cao (Flavell 1982). Thực vật có genome nhỏ và tỉ lệ DNA tự tái bản thấp, thì nó sẽ có chu trình gián phân ngắn và thời gian của một thế hệ cũng ngắn. Một thí dụ điển hình trên *Arabidopsis thaliana*, loài sinh vật được các nhà sinh học phân tử ứng dụng khá phổ biến, bởi vì nó chỉ có 5 tuần cho một thế hệ, nó có hàng loạt các thể mutant, nó dễ dàng trong công nghệ chuyển nạp gen với Ti plasmid, hàm lượng DNA thấp.

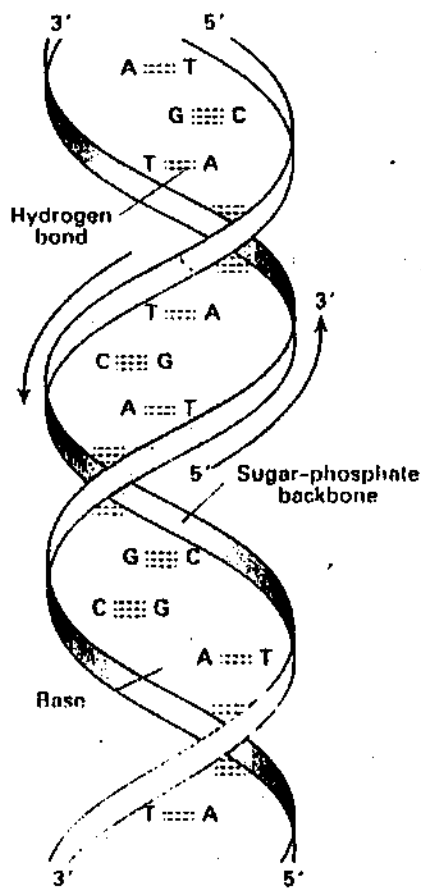
1.5. CẤU TRÚC CỦA CHUỖI MÃ DNA

Có nhiều phương pháp nghiên cứu về DNA của thực vật, nhưng có lẽ phương pháp nghiên cứu động thái của sự tái phối hợp DNA (reassociation) do Flavell (1982) được xem như có giá trị nhất cho phép chúng ta hiểu được cấu trúc genome và được phác thảo đầu tiên.

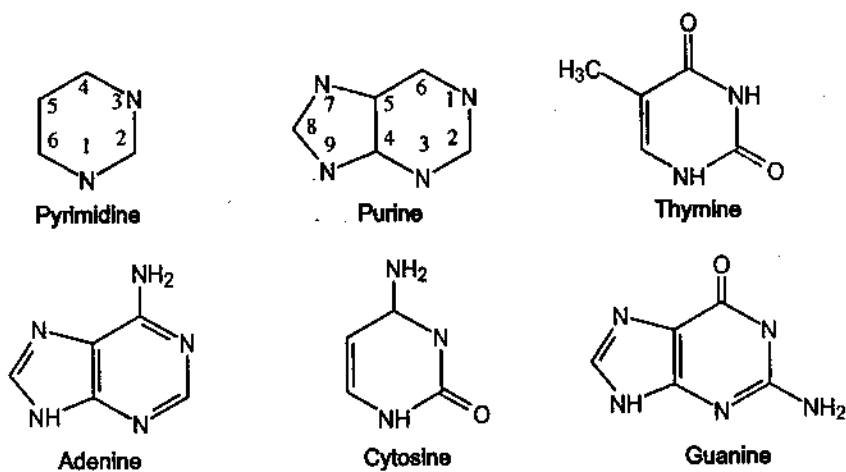
Cấu trúc hóa học của DNA bao gồm :deoxyribose, thymine [T], adenine [A], guanine [G], cytosine [C] và phosphate. Adenine và guanine được xếp vào nhóm base "purine", thymine và cytosine thuộc nhóm base "pyrimidine".

Cấu trúc base + đường được gọi là nucleoside.

Cấu trúc base + đường + phosphate được gọi là nucleotide.



Hình 1.5 : Phân tử DNA dây đôi



Hình 1.6 : Cấu trúc của các base trong phân tử DNA. Thymine và cytosine là pyrimidine, adenine và guanine là purine. Pyrimidine gắn với phân tử đường tại N, purine gắn với đường tại N₉,

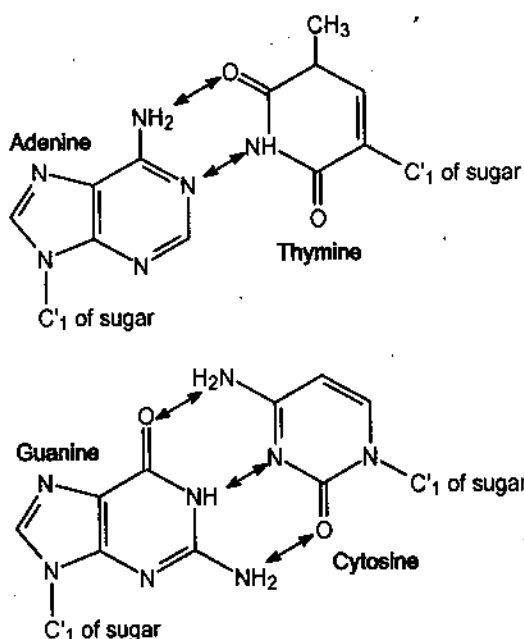
Bảng 1.2 : Cách gọi tên cấu trúc DNA

Thành phần	Tên hóa học	Viết tắt
Base	Adenine	A
	Thymine	T
	Guanine	C
	Cytosine	G
Base + đường	Deoxyadenosine	
	Deoxythymidine	
	Deoxyguanosine	
	Deoxycytidine	
Base + đường + phosphate	Deoxyadenosine -5'- monophosphate	dAMP
	Deoxythymidine -5'- monophosphate	dTMP
	Deoxyguanosine -5'- monophosphate	dGMP
	Deoxycytidine -5'- monophosphate	dCMP
Base + đường + 3 phosphate	Deoxyadenosine -5'- triphosphate	dATP
	Deoxythymidine -5'- triphosphate	dTTP
	Deoxyguanosine -5'- triphosphate	dGTP
	Deoxycytidine -5'- triphosphate	dCTP

DNA là polymer của những nucleotide. Khi hai nucleotide nối với nhau, phân tử vừa được hình thành được gọi là dinucleotide. Tương tự như vậy, 3 nucleotide nối với nhau sẽ được gọi là trinucleotide, bốn gọi là tetranucleotide. Nhiều nucleotide kết hợp với nhau hình thành oligonucleotide. Những nucleotide được nối với nhau bằng liên kết phosphate 5'-3'. Cầu nối phosphate này là cầu nối ester đồng hóa trị, còn gọi là cầu nối phosphodiester. Gốc phosphate còn lại (PO_4^{3-}) dọc trên polymer làm cho nó có tính acid, vì vậy người ta còn gọi nó với thuật ngữ "nucleic acid". Do những cầu nối phosphodiester, đại phân tử polynucleotide có cấu trúc phân cực tại vị trí 3'-hydroxy (3'-OH) và 5'-phosphate (5'-P), ở đầu và cuối bất cứ một nucleic acid nào. Cấu trúc 5' - 3' như vậy còn được gọi là "**palindromic**".

Năm 1953, Watson và Crick đã khám phá DNA là một cấu trúc xoắn đôi. Nó thể hiện giống nhau ở mọi cấu trúc DNA của các sinh vật, trừ một vài sinh vật bậc thấp như virus. Hai dây đơn DNA gắn với nhau bằng cầu nối hydrogen, tạo ra cấu trúc xoắn đôi. A gắn với T bằng hai cầu nối hydrogen. C gắn với G bằng ba cầu nối hydrogen. Cầu nối

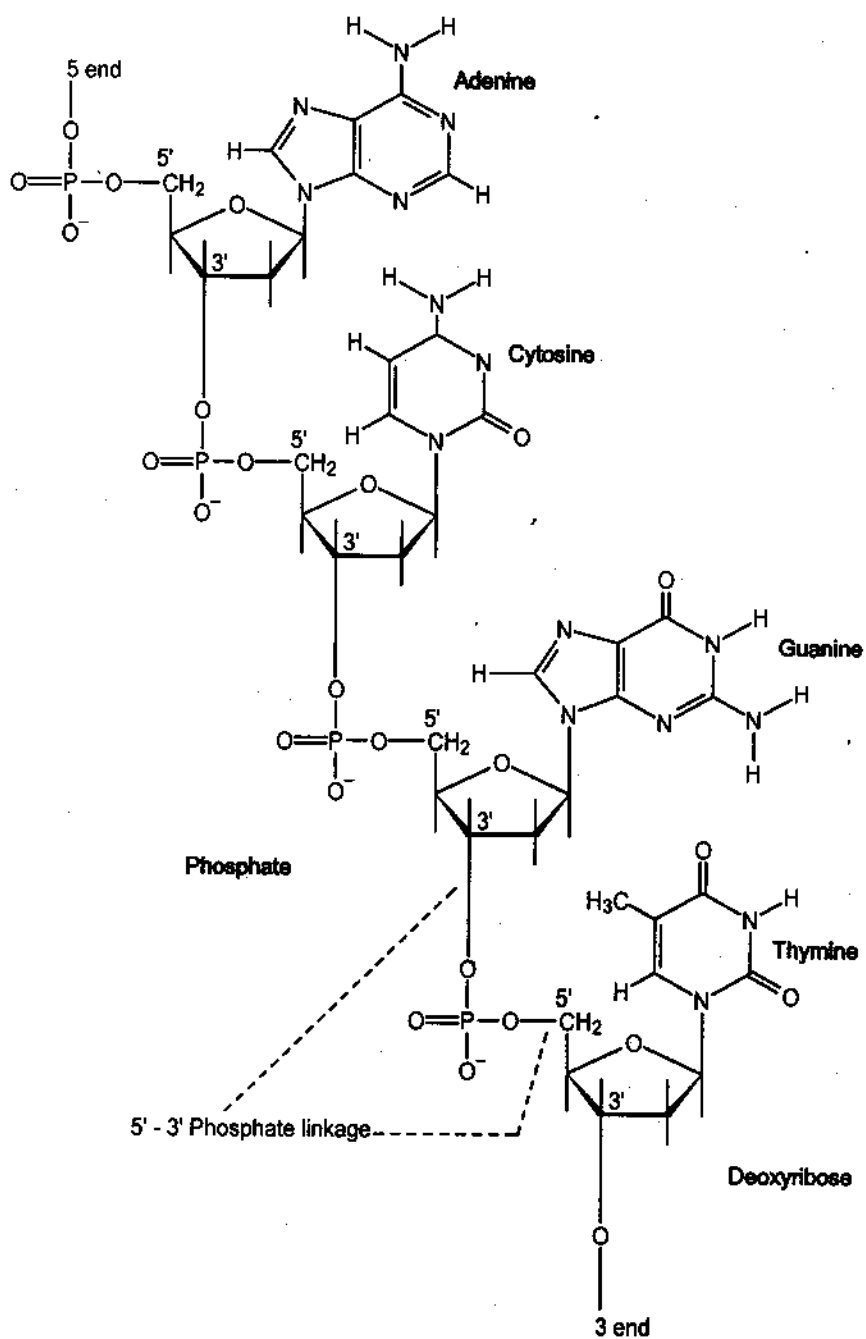
hydrogen này là cầu nối rất dễ bị phá vỡ, làm hai dây đơn của phân tử DNA tách ra. Tiến trình này được gọi là sự biến chất (denaturation hay melting). Tiến trình ngược lại được gọi là sự hoàn nguyên (renaturation hay reannealing). Phương pháp thường được sử dụng để tách hai dây đơn ra là "nhiệt độ" [làm gia tăng sự vận động của các phân tử], và sử dụng alkaline hoặc ure [làm thúc đẩy sự thành lập các cầu nối hydrogen]. Sodium và một số các loại muối có thể giúp cho các xoắn đôi DNA ổn định bằng cách trung tính hóa các gốc phosphate. Tiến trình tách đôi DNA có thể được kiểm soát bằng nhiều phương pháp. Cách đơn giản và hiệu quả nhất là xem xét sự thay đổi của mật độ quang (optical density). Những nucleotide hấp thụ tia cực tím với độ dài sóng tối đa 260nm. Trong quá trình tách đôi dây, giá trị A₂₆₀ sẽ tăng nhiều hơn trên dây đơn so với dây đôi. Người ta sẽ có được một đồ thị giữa nhiệt độ và sự hấp phụ quang phổ. Trong đó, điểm giữa khoảng nhiệt độ trong quá trình tách đôi dây DNA được gọi là **T_m** (melting temperature). Hình dạng của đường biểu diễn giống nhau, nhưng T_m thay đổi tùy theo thành phần base của DNA và một vài yếu tố khác.



Hình 1.7 : A bắt cặp với T bằng hai cầu nối hydrogen, C bắt cặp với G bằng ba cầu nối hydrogen.
Trong RNA Uracil sẽ thay thế Thymine

Dây đơn DNA sẽ không hoàn nguyên một cách hoàn toàn, nếu nhiệt độ giảm quá nhanh như trường hợp đặt mẫu DNA vào nước đá. Người ta lợi dụng đặc tính này để chuẩn bị làm "DNA template" trong việc đánh dấu DNA và trong việc lai DNA / DNA.

Người ta sử dụng thuật ngữ "sequence" để chỉ một chuỗi được mã hóa về mặt di truyền. Chuỗi mã DNA có tính tự tái bản cao chiếm từ 0-10% trong genome của nhóm sinh vật bậc cao (eukaryote). Trong hầu hết các trường hợp, đại phân tử DNA này có một đoạn ngắn các base được lặp lại nhiều lần, tạo một dãy từ 10⁴ đến 10⁷ trên mỗi genome. =

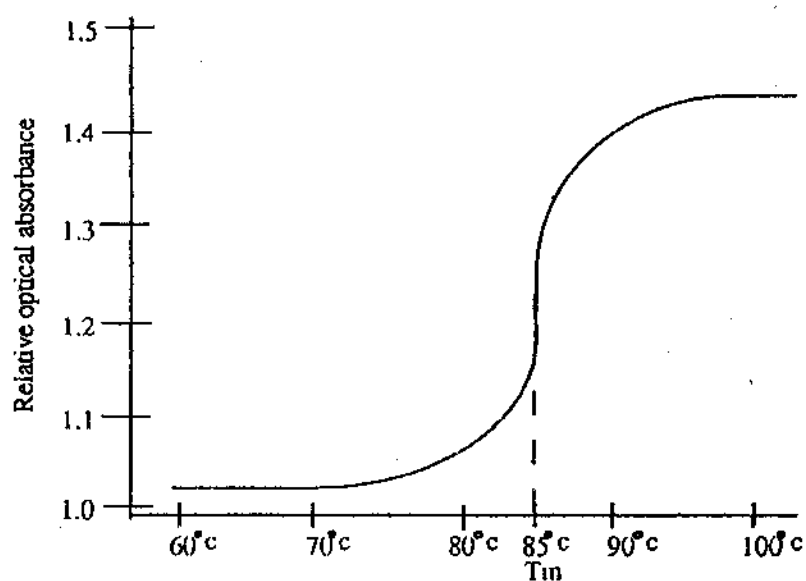


Hình 1.8 : Cấu trúc phân tử DNA với các hợp tử của nó là base + đường deoxyribose + phosphate

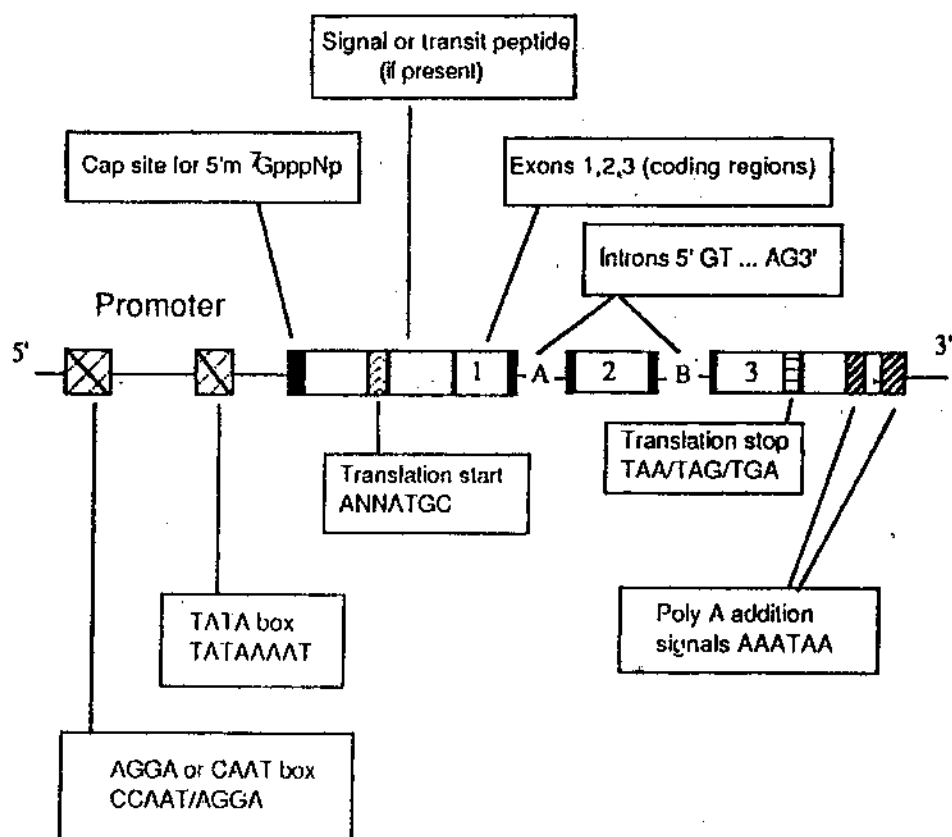
Bảng 1.3 : Chuỗi mã gen tổng quát của cây trồng và so sánh với các nguyên tố tương tự từ Amy3D

TATA box	consensus	TATATATA
	<i>Amy3D</i>	TATATATG
Vị trí chuyển mã	consensus	CTCATC
	<i>Amy3D</i>	ATCATC
Vị trí giải mã	consensus	TAAACAATGGCT
	<i>Amy3D</i>	ACAGATATGAAG
Vị trí phân cắt 5'	consensus	AG/GTAAG
	<i>Amy3D</i>	AG/GTACG
		TC/GTAAG
Vị trí phân cắt 3'	consensus	TGCAG/
	<i>Amy3D</i>	GGCAG/
Dấu hiệu polyadenylation	consensus	AATAAA
	<i>Amy3D</i>	AATAAG
		AGTAAA

Lấy cấu trúc tổng quát của một gen cây lúa làm thí dụ. Gen khởi đầu bằng một promoter region, kế đến vùng mang mật mã di truyền, và kết thúc bằng một terminator (hình 1-10). Promoter region gồm một chuỗi mã DNA tổng quát, và quan trọng hơn cả là các nguyên tố có tính chất điều chỉnh (regulatory). Một chuỗi mã TATAAAT được gọi là "TATA box", định vị tại upstream 3', với khoảng 30 vị trí base từ vị trí đầu tiên giải mã. Bên cạnh TATA box, những nguyên tố điều chỉnh khác cũng được tìm thấy ở vị trí 5' và được gọi là "enhancer" (chất thúc đẩy). Vùng khởi đầu ở vị trí giải mã đầu tiên và vùng kết thúc giải mã sẽ được chuyển hóa thành RNA, nghĩa là đường ribose sẽ thay thế đường deoxyribose, và base Uridine [U] thay cho base Thymine [T]. Vùng giải mã gồm có: một chuỗi không giải mã (untranslated) ở cả vị trí 5' và 3', một chuỗi khởi đầu giải mã, thể exon [mang mã], thể intron [không mang mã], và chuỗi mã đánh dấu bổ sung poly A [còn gọi là đuôi poly A]. Trước khi thể exon có thể được sử dụng để giải mã nhằm sản xuất ra protein/enzyme, thì RNA cần được chuyển thành mRNA [RNA thông tin]. Có hai sự kiện quan trọng làm chuyển đổi intron và đuôi poly A. Chuỗi mã vùng tiếp cận của intron trong RNA thường là GU ... AG (để dễ nhớ chúng ta dùng từ get up and go). Dấu hiệu của polyadenylation được đặt sau codon giải mã cuối cùng. Thí dụ gen cây lúa mã hóa bằng alpha-amylase ở hình 1-11. Đặc điểm của gen này được mô tả trong bảng 1-3.



Hình 1.9 : Nhiệt độ T_m trong hiện tượng thoái hóa dây DNA [denaturation]



Hình 1.10 : Cấu trúc tổng quát của một gen thực vật

10 30 50 70 90
 GATCTTCACACCTGTGCTAGCTACTCCACTGCTCCATAGGCAATCATCAATCAGTAATCCGTTCTGAAAGAAGATATAGGTGTGCGGCAATCAGGAAC
 110 130 150 170 190
 GTTCTAGTTCGTCTAGAAATCAGCAGCTGCTAAGTTAGCATCTCGATGAACTTAAATGCTGCTGCGGCGGTCCGCGGAGATGAAGTTTGTGATAAAC
 210 230 250 270 290
 TTGGTCATCAGCATTCATATATGTCCTGGTCTACGGAGTAGTTTCATCAGCAACATACACCTACTTCTACTCTATCCATTGGATTGCTCATGCGCGCTT
 310 330 350 370 390
 TGATATGGAATTTGTAATGAACCTTGGTTATGACTTATGACATACGATACCTCGTAACATTATAGATATCTGACATAAATTCATCAACTACAATAGATGAG
 410 430 450 470 490
 ATGGGTAGTCTTAGTAGAACAGTAGTCTCTCTTTCGGCTTGGTCCATTGGCTGATGACGATGAACAACTCGGACTCATTGATTCCAGCATTATCTGATT
 510 530 550 570 590
 CTCGCAATTCGAGGTCCGGATTAGGCTCTACCGAGATGTGGATAGCAATTCGCAATGTGAGGAATGAAGGAGGACGAGCCATATGTCGATATACATGACG
 610 630 650 670 690
 GGAGATCAAGCGCCAGTCAAGAGCCTAAGTCAACCCCTATTATATACGATCAGCCTGCTAGAACACCTAGCACTGTCTTTTTTGTCTGAACCTCTGAAGA
 710 730 750 770 790
 TGAAAGGCTCAGAGAAATGGCTCGCTTATCCAAGCCGGGATGGATGGAGGAGGATAGCCGCGCCCGCTCAGGCAGTCTGTCGCGATCACGCCGCC
 810 830 850 870 890
 GCATCCCGTTCGCTTGGAGACCGGCGCCGAGCGCGCGGCGGCTACGTCGGCATGCTTTATTTGCTTATCCATATCCAGCCCATTTATTTGGT
 910 930 950 970 990
 CGTCTCTCTGATCATTTCTATTCCCTCGCCACGGTGACCGTGCCTCCCGGCTGCTTATATATGCCCCCGAGCTCGAGGTCTATCCGACGAAACACATCG
 1010 1030 1050 1070 1090
 ATCATCTCATCTACAGAGATCGATCAGTAGTGGTTAGCAGCAACTCACTATCGAACACGGTTTCAGCTTACACAGATATGAAGAACACCAAGCAGCTT
 1110 1130 1150 1170 1190
 5'CATCTACAAGAGATCGATCAGTAGTGG 3' M K N T S S L
 GTGTTTGTCTCTCTGCTGCTCTGCTCTGACCTTGAACCTGAACTCGGCTCAAGCACAGGCTCTTCCAGGTACGTAGTACTCTACTACCCATCACTTTC
 C L L L L V V L C S L T C N S G Q A Q V L F Q
 1210 1230 1250 1270 1290
 TGTGAAGACTTTTGTGAAGAACACATTAGAAATTTGAGATATTTATGTGTGATCGATTGATCACTTACCTACTTATAACATGCATCGACGGGTTTC
 G F
 3' TTGACCTTCAGCACTCTCTCTGTTCC 5' 1350 1370 1390
 AACTGGGAGTCTGTGAAGCAGCAGGGTGGCTGGTACAACATGTTGAAGGCCAAGTCGACGACATCGCCAGGCCGGGGTCAACCCAGCTCTGGCTGCCGC
 N W E S W K Q Q G G Y N M L K G Q V D D I A K A G V T H V W L P P
 1410 1430 1450 1470 1490
 CGCGCTCGCACTCCGTGCGCGCCGAGGGGTACATCGCGGGGCTCTTACGACCTGACGCGTCCAAGTACGGCACGGCGGGGAGCTCAAGTCCCTGAT
 P S H S V A P O C Y N P G R L Y D L D A S K Y G T A A E L K S L I
 1510 1530 1550 1570 1590
 CGCGGCTTCCAGCGGAAGGGGCTCAGTGGCTCGCGGACGCTGATCAACACCGGTCGCGCGAGAGAAGGACGCGCGCGGCTGTACTCGGTGTTT
 A A F H G K G V Q C V A D V V I N H R C A E K K D A R G V Y C V F
 1610 1630 1650 1670 1690
 GAGGGCGGACGCGCCGACCCGCTCGACTCGGGCGCGGCTGATCTCAGCGACGACGACGCGCTCTCGACGCGACGCGGACCGGACACCGCGGAGG
 E G G T P O R L D W G P G M I C S D D T Q Y S D G T G H R D T G E G
 1710 1730 1750 1770 1790
 GGTTCGGCGCGCGCGGACATGACCACTCAACCGCGGCTCAAGCGGAGCTCAGCGACTGGCTCAACTGGCTCAAGTCCGACGCTCGGCTTCGACGG
 F G A A P D I D H L N P R V O R E L T D N L M W L K S D V G F D G
 1810 1830 1850 1870 1890
 CTGGCGCTTCGACTTCGCGCAAGGATACTCCACGGACATCGCTAAGATCTACGTCGAGAGCTGCAAGCGCGGCTTCGCTCGTCCGCGAGATATGGAACCTCG
 W R L D F A K G Y S T D I A K H Y V E S C K P G F V V A B I W N S
 1910 1930 1950 1970 1990
 CTGAGCTACAAGCGCGACGCGAAGCGCGCGCCCAACAGGACGAGGCGCGGCGAGGCTCGGTGAACCTGGGTCAACCGCGCTCGCGCGCGCGCGGATGACCT
 L S Y N G D G K P A A N Q D O G R Q E L V N M V N A V G G P A M T F
 2010 2030 2050 2070 2090
 TCGACTTCACCAAGGCGCTCTCTCAGCGCGCGCTCCAGCGCGGCTGTGGCGGCTCGCGGACGCGCAAGCGCGCGCGCGCGCGCTCGGCTGGCT
 D F T L K G L L Q A G V Q G E L W R L D G K G K A P G H I G N L
 2110 2130 2150 2170 2190
 GCCAGAGAGCGCGCTCAGCTTCGTCGACACACGACCGCGCTCGACGCGAGAGCTTTGCGCGTTCGCGCTCCGACAGGCTCATGACGGCTACGCGCTAC
 P E K A V T F V D N H D T G S T Q K L M P F P S D K V H Q G G Y A Y
 2210 2230 2250 2270 2290
 ATCCTCACCCACCCCGAGTCCCGCTGCATCGTAAGCAACCATGCATTATAGTATTATACCATGTCTGATTAACTCCACCGTACAGCTGTCTCGAT
 I L T H P G V P C I
 2310 2330 2350 2370 2390
 GAACGCTTCTTGTGCGAGTTCTACGACCATATGCTGAGCTGAACTCAAGCAGGAGATAACCGGCTCGCGGCGGATCAGGAGAGGAACGGCATCAACG
 F Y D H M F D W N L K Q E I T A L A I R E R N G I N A
 2410 2430 2450 2470 2490
 CCGGAGCAGCTCCGAGTCTCTCCCGACCGGACGATACGTCGCGCTGCTGACGAGAGAGTCTGGTGAAGATCGGGACGAGGTACGAGCTCGGG
 G S K L R I V V A D A D A Y V A V V D E K V M V K I G T R Y D V G
 2510 2530 2550 2570 2590
 CAACGCGGTGCGCTCGGATTTCCATCAGACGGTGCACGGCAAGGACTACAGCGTCTGGGAGAGGGGTCCTCCGCGTCCCGCGGGCGGCGACCTATAG
 H A V P S D F H O T V H G K D Y S V W E K C S L R V P A G R H L
 2610 2630 2650 2670 2690
 CCGGCTCAAGCCCTAAACTGAACGGUATAGTCATCTCAACAGCTTCTACCGGCAAGAATTACGATTTCTTACTTTTTCAGTCAATTAAATTAT
 2710 2730 2750 2770 2790
 GGTTTTATATATGTAATTTTGTATCCGATTGTAGCGTTCCGATAAGTACGAGGCTCTCTAGCCTCTAGGTTAATTGCGGGCATATGTAGCTTCCAG
 2810 2830 2850 2870 2890
 TTAATTGCTTTTATATCAACAGCTTTGTAATCTTCTGCAATATATAATCTCAGCTTCAAGATGCTATAAAAAATCATACTGCACCGATCACTGAGTTT
 2910 2930
 TTATATACTGGTTTAAAAAGTGAGCACAACTACTAATT

Hình 1.11 : Chuỗi mã DNA của gen alpha amylase trong cây lúa Amy3D. Vùng mang mã
 di truyền được xác định bởi sequence dây thẳng đối với cDNA của nó, pOS137.
 Sequence của amino acid là từng chữ in ở hàng dưới, sau mỗi ba nucleic acid
 (codon). TATA box, CATC box là dấu hiệu polyadenylation được gạch dưới.

1.6. DNA VỆ TINH (SATELLITE DNA)

Một vài DNA vệ tinh của động vật được tìm thấy đầu tiên chứng minh rằng có những chuỗi mã giản đơn, có tính tự tái bản, gắn ở tâm động (centromere) hoặc ở vùng dị nhiễm của nhiễm thể. Tuy nhiên thuật ngữ "satellite" có thể được sử dụng cho bất cứ DNA sequence nào được phân lập về mặt vật lý, nhờ thay đổi nồng độ isopycnic trong dung dịch CsCl, mà không ngụ ý như một ori đặc biệt hay một chức năng đặc biệt nào đó. Tỷ trọng của DNA tùy thuộc vào bản chất của muối, ở đó DNA đang ở trạng thái dây đơn hay dây đôi, và tỷ trọng này bị ảnh hưởng bởi pH, sự thêm vào các kim loại nặng hoặc thuốc nhuộm DNA, và những thuốc khác. Trong môi trường CsCl, yếu tố chính xác định tỷ trọng là thành phần của base, trong khi đó trong Cs₂SO₄, actinomycin D và Hg²⁺ gắn với sự thay đổi tỷ trọng của DNA giàu base GC, và Ag⁺ gắn với DNA giàu AT. Nhiều vệ tinh được mô tả trong thực vật bậc cao (Ingle và ctv. 1973). Nó có một dây dài DNA chứa những chuỗi mã có tính tự tái bản. Nó có thể đại diện cho một phần nhỏ hoặc lớn của toàn bộ phân tử DNA. Sự phát hiện của vệ tinh như vậy tùy thuộc vào mức độ chính xác của phương pháp sử dụng. Một loại vệ tinh nào đó được định tính, thì chuỗi mã của nó cũng như vị trí của nó trên nhiễm thể sẽ được xác định bằng kỹ thuật lai *in-situ*. Chiều dài có tính lặp lại nhiều lần này của vệ tinh thay đổi từ vài bp đến hàng trăm bp. Người ta hi vọng các genome ở cơ quan thực vật có thể được bù đắp thiếu sót của nó bằng DNA vệ tinh của nhân.

1.7. CHỨC NĂNG CỦA CHUỖI MÃ DNA CÓ TÍNH LẬP LẠI

Người ta thấy rất rõ có một vài chuỗi mã DNA có tính lặp lại ở trong vùng kiểm soát, giữ chức năng điều tiết việc chuyển mã. Những sequence này có rải rác trong DNA dạng single copy, và trong DNA lặp lại có tính chuyển mã như là những gen rRNA. Tín hiệu bắt đầu và kết thúc sự chuyển mã thì tương đối ngắn, trong khi các chuỗi mã khác có thể còn có sự tương tác phức tạp, sự ức chế, hoặc kích hoạt vùng chứa DNA.

1.7.1. Chuỗi mã có tính chất lặp lại cao

Chuỗi mã DNA có tính chất lặp lại cao chiếm từ 0-10% của eukaryotic genome. Trong hầu hết các trường hợp, đại phân tử DNA này có một đoạn ngắn các base được lặp đi lặp lại nhiều lần, tạo một dây từ 104 đến 107 trên mỗi genome. Vì đoạn có tính chất lặp lại cao này thường có thành phần base khác với DNA trong eukaryotic genome, cho nên chuỗi mã DNA có tính chất lặp lại cao thường tạo ra dây băng ở một vị trí vệ tinh nào đó, với một băng chính theo một CsCl gradient. Do đó, chuỗi mã DNA có tính chất lặp lại cao được gọi là DNA vệ tinh [satellite DNA].

Đơn vị của một lần lặp lại có thể rất ngắn, chừng hai nucleotide [gọi là dinucleotide repeat]. Một chuỗi mã có chứa hai hoặc một vài nucleotide repeat như vậy, được gọi là sự tái lập giản đơn của chuỗi hay microsatellite. Thí dụ như: một microsatellite định vị gần gen *Adh-2* có 9 repeats của GA và 4 repeats của GAT (Wu và Tanksley 1993) [bảng 1-4].

Khi số đơn vị lặp lại khá nhiều, người ta sử dụng thuật ngữ "minisatellite". Thí dụ: chuỗi mã AGGAAGGGAGGA là minisatellite đã được tìm thấy trên cây lúa. Minisatellite repeat này có trong hàng nghìn bản sao của genome cây lúa.

Bảng 1.4 : Những repeat đơn giản của chuỗi mã trong gen cây lúa

Coumarate-CoA ligase	(GT) 8
Type II light harvestin	(GT) 8
chlorophyll a, b binding protein	
<i>Adh2</i> gene	(GA) 9 (GAT) 4
<i>rcbs</i> gene	(GT) 3 (AAG) 7
<i>rgpl</i>	(GAA) 7
<i>Oryza cystatin</i> gene	(AT) 40
<i>Phyt 18</i> gene	(AT) 21

1.7.2. Những chuỗi mã có tính chất lặp lại trung bình

Chuỗi mã có tính chất lặp lại trung bình chỉ khoảng 40% trong genome cây lúa và nó chứa các chuỗi mã DNA được lặp lại từ vài lần đến hàng trăm lần. Chuỗi mã có tính chất lặp lại trung bình có tính chất dị hợp nhiều hơn chuỗi mã có tính lặp lại cao. Trong phân tích Southern blot, nó tạo ra từ nhiều vạch đến một đốm. Chuỗi mã DNA có tính chất lặp lại trung bình có xu hướng phân tán một cách điển hình trong genome hơn là tập hợp thành nhóm một cách ngẫu nhiên.

1.7.3. Chuỗi mã đơn có tính chất sao chép

A	SC	SC	SC	SC			
B	SC	R	SC	R	SC	SC	R
C	MR	HR	MR	SC	MR	HR	

Hình 1.12 : Sắp xếp của chuỗi mã tùy theo các dạng khác nhau của DNA

[A] chuỗi mã đơn có tính sao chép không gián đoạn (SC)

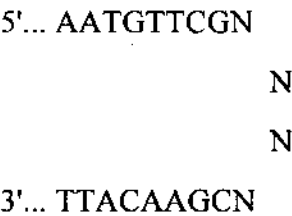
[B] chuỗi mã đơn xen lẫn chuỗi mã có tính lặp lại (R)

[C] chuỗi mã có tính lặp lại cao (HR) và trung bình (MR) xen lẫn nhau

Chuỗi mã đơn có tính chất sao chép chiếm khoảng 60% DNA của cây lúa và hầu hết các gen đều thể hiện tính chất này. Thí dụ *Amy3D* gene [Hình 1-11]. Khi DNA từ đoạn này được sử dụng như là chất thăm dò trong lai DNA, trong kỹ thuật Southern blot, thì người ta có thể quan sát các dãy đơn hoặc dãy đôi [couple bands]. Giống như DNA có tính chất lặp lại trên, chuỗi mã đơn DNA có tính chất sao chép cũng phân tán giữa những đoạn lặp lại. Hình 1-12 cho thấy vị trí tương đối của chuỗi mã đơn, chuỗi mã trung bình, và chuỗi mã lặp lại cao.

1.7.4. Lặp lại có tính chất đảo đoạn

Lặp lại có tính chất đảo đoạn [inverted reverted] gồm có những đoạn DNA bổ sung nhau cho từng quãng. Thí dụ : 5'... AATGTTTCGNNNGAACATT... 3' trong đó N là ký hiệu cho bất cứ nucleotide nào. Hai vùng cách nhau bởi NNNN được ký hiệu là R(AATGTTTCG) và R'(CGAACATT). Hai vùng R và R' có tính chất bổ sung nhau, do đó khi dây DNA được đặt trong điều kiện tái phối hợp trở lại [reassociation], nó sẽ nhanh chóng gấp lại, tạo thành dạng hairpin.



Nếu chuỗi mã của NNNN vừa đủ dài, chúng ta sẽ có một thân chứa các inverted repeat, cộng thêm một hình vòng [loop] dính ở giữa chúng.

Thí dụ cấu trúc stem-loop của cây lúa trong hình 1.4. Chuỗi mã DNA có từ một intron của gen alpha-amylase. Chính cấu trúc stem-loop có thể gây ra nhiều vấn đề trong kỹ thuật khuếch đại PCR sau này.

Cấu trúc stem-loop có đặc tính tương tự như những nguyên tố chuyển mã bởi vì hầu hết ở các stem-loop được tìm thấy, người ta cũng phát hiện ra tính chất lặp lại trực tiếp [hình 1.4]. Sự nhảy vọt [jumping] của những nguyên tố chuyển mã này có thể gây ra nhiều đột biến DNA, bao gồm tính chất cộng thêm, thiếu, chuyển vị, v.v...

Các loài thực vật có cùng một genus, thường khác nhau về DNA vệ tinh có tính lặp lại. Người ta có thể nói rằng : không phải tất cả các chuỗi mã có tính lặp lại đều có một chức năng chuyên môn nào đó. Do đó, chúng ta cần phải nghiên cứu riêng từng trường hợp một cách thận trọng.

1.8. GEN MÃ HÓA TRONG TỔNG HỢP PROTEIN

Bằng nhiều cách khác nhau người ta ước đoán có khoảng 40.000 đến 100.000 gen ở

thực vật bậc cao, là những gen đại diện cho mã hóa protein (coding). Số lượng này lớn gấp 10-25 lần so với *E. coli*. Một gen trung bình coding cho một protein có 350 amino acid, sẽ cần ít nhất là 1050 bp. Nếu kể cả intron và vùng kiểm soát, thì nó có thể tăng lên 2000 bp, như vậy cũng còn rất nhỏ bé so với một vài gen của thực vật khác. Sử dụng số liệu này, người ta phải cần đến 80-200 triệu bp để tạo mã cho số gen được dự đoán có tính chất cơ bản. Áp dụng phương pháp động thái học trên DNA của *Arabidopsis thaliana*, người ta thấy rằng genome này là 70 triệu bp (Leutwiler và ctv. 1984). Điều này có thể xét chỉ có 35.000 gen của nó được thảo luận trước đây, có nghĩa là gen của *Arabidopsis thaliana* rất lý tưởng cho nghiên cứu (thể intron rất ít và rất ngắn) so với cây khác có genome lớn hơn.

Nhiều gen chứa các chuỗi intron xen kẽ làm cản trở việc mã hóa. Có những nhóm của các sequence có liên quan, với sự khác nhau về vùng mang mật mã protein, thể intron, vùng không giải mã của mRNA, hoặc vùng không giải mã có tính chất upstream hay downstream của nhiễm thể. Phân tích sequence của những nhóm đa gen như vậy có thể cung cấp cho chúng ta biết được sự tiến hóa và sự điều tiết trong thể hiện gen.

Gen điều khiển protein dự trữ alcohol-soluble prolamin của cây bắp là một trong những thành viên của nhóm đa gen này (large multigene family). Những gen này không có intron và mã hóa thành một nhóm protein trong phôi nhũ hạt, giàu glutamine (z1 zein), hay proline, hay cysteine, hay methionine (z2 zein). Nhóm z1 hình thành một subfamily. Có khoảng 25 gen đối với zein z1A, 20 gen đối với z1B, 15 đối với z1C, và 5 đối với z1D. Những gen này đã được phân tích trên bản đồ, với nhiễm thể mà nó gắn vào bằng phân tích liên kết gen với marker. Các gen z1A định vị trên nhiễm thể số 4, 7 và 10. Các gen của z1B định vị trên nhiễm thể số 4 và 7. Các gen z1C định vị trên nhiễm thể số 4. Các gen z1 xuất hiện theo kiểu cluster (chùm, nhóm) khi lập lại trên cùng một nhiễm thể. Trong khi các gen z2 có số lượng ít hơn, và có một subfamily nhỏ hơn. Các gen zein thể hiện một cách đặc hữu trong quá trình dự trữ zein ở dưới dạng protein dự trữ (Heidecker và Messing 1986).

Họ gen z1 được xem như tăng thêm nhờ tính chất lặp lại của sequence và sự quần chéo không cân đối. Cấu trúc của gen này bao gồm một vùng mang mã peptide tín hiệu có khả năng bảo quản cao của 21 amino acid, gốc cuối (termini) N- của 36-47 amino acid và C- của 10 amino acid. Phần giữa trung tâm của protein này chứa 6-10 đoạn lặp lại của một sequence nguyên thủy (ancestral) mã hóa hàng loạt các glutamine residues, cộng với những amino acid khác như leucine, proline và alanine.

Một subunit của rubisco và diệp lục tố a/b gắn với protein (*Cab*) được xem như thí dụ điển hình về protein được mã hóa bằng multigene family. Tám gen ở subunit của rubisco đã được tìm thấy trên cây *Petunia*. Sự chuyển mã của một gen cho thấy ước khoảng 50% subunit của mRNA tổng số được tăng lên. Sự điều tiết việc thể hiện gen của subunit thường phối hợp với DNA sequence đến vị trí 5' của vùng mang mã. Tuy nhiên, phân tích gen của *petunia* cho thấy sequence đến vị trí 3' của vùng mã protein cũng cho một sự thể hiện rất

cao. Trong bắp có 6 gen điều khiển Cab protein. Ba gen thể hiện ở bó mạch của bẹ lá và tế bào diệp lục (mesophyll). Nghiên cứu sâu hơn về sequence của gen giúp chúng ta giải thích sự thể hiện gen rất khác nhau trong tự nhiên. Không có tất cả các member của một họ đa gen đều được thể hiện. Thí dụ như, chỉ có một nửa của gen *z1* tỏ ra hoạt động. Đối với một vài họ gen khác, nó được biết nhờ sự có mặt của những gen giả (pseudogenes). Những gen này giống như gen bình thường trong một vài trường hợp, nhưng nó không đầy đủ, thiếu sequence kiểm soát chuyển mã, hoặc chứa những stop codon. Thí dụ pseudogene actin của khoai tây không có intron, nhưng có đuôi poly A, cho thấy nó có thể gia tăng nhờ sao lại mRNA bằng men reverse transcriptase. Vùng mang mã chứa 2 stop codon. Có một repeat trực tiếp mang 18 bp ở mỗi đầu, gắn vào chuỗi mã trong genome khoai tây (Drouin và Dover 1987)

1.9. CÁC NGUYÊN TỐ CHUYỂN VỊ (Transposable)

Các nguyên tố chuyển vị, nguyên tố kiểm soát, hoặc nguyên tố gắn vào đều là những chuỗi mã DNA, nó có thể được đính vào nhiều vị trí khác nhau trên nhiễm thể. Một vài nguyên tố có khả năng di chuyển xung quanh genome, cũng như xoay quanh chính nó, hoặc chuyển động dưới ảnh hưởng của những chuỗi mã khác (Nevers và ctv. 1986). Trong việc này, đôi khi người ta còn gọi nó với thuật ngữ "jumping genes". Sự chuyển động của nó có thể được tìm thấy về mặt di truyền, bởi vì một nguyên tố di động đã được gắn vào một gen cấu trúc, gen này cũng có thể bất hoạt (inactivated). Theo trình tự, việc gắn vào của một nguyên tố chuyển đổi với vùng kiểm soát, có thể làm thay đổi sự thể hiện của gen kế cận. Cũng có trường hợp một nguyên tố chuyển đổi bị tách ra khỏi một gen không hoạt động, gen này lập tức trở lại hoạt động.

Những nguyên tố chuyển vị trên cây bắp đã được tìm thấy và được nghiên cứu về đặc tính chi tiết do McClintock (1948, 1951), trước khi cấu trúc DNA được công bố. Những nguyên tố di truyền như vậy được tìm thấy và được định tính sau đó trên vi khuẩn, men (yeast), động vật không xương sống, động vật có xương sống, và thực vật khác. Nhiều nguyên tố đã được cloned hóa và được viết ra chuỗi mã (sequencing). Công trình nghiên cứu của McClintock sau hơn 40 năm về nguyên tố chuyển vị (transposable elements) trong cây bắp, đã được nhận giải thưởng Nobel về Sinh lý và Y học.

Một loại hình về nguyên tố gắn vào (insertion element) trên cây bắp, là Dissociation (*Ds*) không có thể di chuyển do chính nó, nhưng nếu có một nguyên tố khác là Activator (*Ac*) thì nó có thể di chuyển đến một vị trí di truyền mới. Người ta đã tìm thấy ảnh hưởng kiểu hình của những nguyên tố này, khi người ta làm đứt đoạn gen đặc biệt nào đó như gen *shrunk* (làm nhăn hạt bắp), gen này ảnh hưởng đến enzym có tên gọi là "sucrose synthase" trong phôi nhũ hạt. Người ta đã chạy sequence một vùng có độ lớn 4,2 kp (1kp = 1000 bp), để làm cho tương ứng với *Ds*, nó tạo ra một phần của sequence gắn vào locus của

gen shrunken. Chuỗi mã của *Ds* có 2040 bp. Chuỗi mã này có một thể lặp lại ngược (inverted repeat) với độ lớn 11 bp (TAGGGATGAAA) nằm ở mỗi đầu. Người ta xem như nó tạo ra một repeat có tính chất trực tiếp, ngắn, trên nhiễm thể.

Nguyên tố *Ac* với 4563 bp cũng đã được chạy sequence. Nó cũng cho thấy có cùng một loại repeat ở đoạn cuối, có tính đối ngược như *Ds* và cũng tạo ra 8 cặp base lặp đoạn ở vị trí dính vào trên nhiễm thể. Kết quả DNA sequencing cho thấy một vài nguyên tố *Ds* có liên hệ gần với *Ac* nhưng chúng luôn có những đứt đoạn hoặc đoạn gắn vào có tính chất luân phiên. Chuỗi mã của *Ac* cho thấy : nó chứa khung mở cho phép đọc (open reading frames viết tắt là ORFs), đó là vùng mã hóa protein sau này. Tiến trình được phát hiện nhờ nghiên cứu hệ men putative RNA polymerase, theo sau đó là nhờ các tín hiệu polyadenylation.

TÀI LIỆU THAM KHẢO

1. Bennett MD, JB Smith, JS Heslop-Harrison. 1982. *Nuclear DNA amount in angiosperm*. Proc. R. Soc. London B, 216: 179-192.
2. Bennett MD, JB Smith. 1976. *Nuclear DNA amounts in angiosperm*. Phil. Trans. R. Soc. London, Ser. B. 274: 227 - 274.
3. Bryant J. 1982. *DNA replication and the cell cycle*. In *Nucleic Acids and Protein In Plants*. II. Structure, Biochemistry and Physiology of Nucleic Acids. Eds. Parthier and D. Boulter, Springer Verlag, Berlin, Heidelberg and New York, 75 - 110.
4. Drouin G, GA Dover. 1987. *A plant process pseudogene*. Nature (London) 328:557-558
5. Flavell RB 1982. *Chromosomal DNA sequences and their organisation*. In *Nucleic Acids and Protein in Plants*. II. Structure, Biochemistry and Physiology of Nucleic Acids. Eds. Parthier and D. Boulter, Springer Verlag, Berlin, Heidelberg and New York, 46-74
6. Grierson D, SN Covey. 1988. *Plant molecular biology*. Blackie et Son Ltd. 3rd Edition. Glasgow and London. P.1-19
7. Heidecker G, J Messing. 1986. *Structural analysis of plant genes*. Ann. Rev. Plant Physiol. 37:439-466
8. Ingle J, GC Pearson, J Sinclair. 1973. *Species distribution and properties of nuclear satellite DNA in higher plants*. Nature New Biol. 242: 193-197
9. McClintock B. 1948. *Mutable loci in maize*. Yerab. Carnegie Inst. Wash. 47:155-169
10. McClintock B. 1951. *Chromosome organization and genetic expression*. Cold Spring Harb. Symp. Quant. Biol. 16:13-47
11. Nevers P, NS Shephend, H Seadler. 1986. *Plant transposable elements*. Adv. Bot. Res. 12:103-203
12. Leutwiler LS, BR Hough Evans, EM Meyerowitz. 1984. *The DNA of Arabidopsis thaliana*. Mol. Gen. Genet. 194:15-23

HOẠT ĐỘNG CỦA CÁC ENZYME TRONG PHÂN TÍCH GENOME

Enzyme là những hoạt chất sinh học rất quan trọng trong công nghệ di truyền, nhất là quá trình cắt và nối DNA. Nó tác động các phản ứng sinh hóa trong tế bào sống, kiểm soát và điều tiết tính hữu hiệu (efficiency) và tính chuyên biệt (specificity).

Tính hữu hiệu ở đây có nghĩa là mức độ ảnh hưởng cao, có phạm vi rộng.

Tính chuyên biệt có nghĩa là ghi nhận ở mức độ phân tử một cách đặc thù giữa enzyme và nucleic acid của nó.

Chức năng xúc tác đáp ứng theo cấu trúc 3 chiều của protein (3-D).

Enzyme là công cụ cần thiết không chỉ trong kỹ thuật "recombinant DNA", mà còn trong nhiều lĩnh vực khác có quan hệ với sự chuyển hóa về mặt hóa sinh. Căn cứ theo những yêu cầu này, ngành học chuyên về enzyme (enzymology) đã công bố nhiều ứng dụng về kỹ thuật, được nhiều nhà nghiên cứu thuộc nhiều lĩnh vực khác nhau sử dụng.

Theo định nghĩa cổ điển, enzyme là một protein có nhiệm vụ xúc tác (catalytic). Là một chất xúc tác, nó chỉ có thể phát triển một phản ứng theo hướng thuận lợi có tính chất nhiệt động học mà thôi. Nó không thay đổi hướng của một phản ứng. Bất cứ một RNA nào cũng được tìm thấy có chức năng xúc tác, đơn độc hoặc phối hợp với protein. Những RNA như vậy được gọi là *ribozymes*. Ribozymes là một dạng mới khá thú vị của enzyme (Eun 1996), dùng trong kỹ thuật "recombinant DNA".

Mặc dù người ta ghi nhận khả năng xúc tác của enzyme phổ biến như vậy, nhưng cơ chế chi tiết của tính chất xúc tác này được nghiên cứu rất ít. Protein và kỹ thuật gen hiện được ứng dụng để cải tiến và thiết kế lại enzyme không chỉ giải thích mối quan hệ giữa chức năng và cấu trúc, mà còn tạo ra những cơ chất mới và phản ứng chuyên tính mới, nhằm cải tiến tính ổn định, cũng như hiệu quả xúc tác. Việc ứng dụng nguyên tắc về enzyme trong sản xuất kháng thể đã mở ra triển vọng tạo ra các kháng thể nhái theo cấu trúc enzyme xúc tác gọi là "abzymes". Những abzymes này đang ở giao điểm của hai ngành học: miễn dịch học và hóa học. Đó là công cụ có tính chất enzyme học vô cùng giá trị.

2.1. THÔNG TIN DI TRUYỀN

Protein là sản phẩm được giải mã của thông tin di truyền, với mã hiệu của sequence RNA hoặc DNA. Mỗi amino acid có 3 nucleotide tương ứng được gọi là codon. Có 64 codon đã được biết dựa trên sự phối hợp của 4 nucleic acid trong tự nhiên, 61 codon chuyên môn hóa amino acid, và 3 codon còn lại làm nhiệm vụ giải mã tín hiệu stop (dừng

lại) [bảng 2.1]. Có 20 amino acid trong tự nhiên làm vật liệu để kiến trúc các sườn (blocks) của protein (hình 2-1). Nhiều amino acid chuyên tính với hơn một codon, hiện tượng này được gọi là "codon degeneracy".

Những sequence của DNA mang mã hiệu cho các chuỗi RNA có tính chức năng hoặc protein thì được gọi là "gen". RNA mang thông tin để giải mã di truyền cho một protein nào đó được gọi là RNA thông tin (mRNA). Nó được bắt nguồn từ DNA theo tiến trình được gọi là chuyển mã (transcription). Dòng thông tin này từ DNA sang RNA, rồi đến protein, mỗi lần như vậy được gọi là "central dogma" của sinh học phân tử, nhưng người ta chưa biết dòng thông tin ngược lại từ RNA đến DNA, tiến trình này có thuật ngữ là "reverse transcription". Trong sinh học, virus thành lập một phương thức đơn giản nhất về thông tin di truyền. Thí dụ, DNA virus thường thể hiện gen của nó theo cách DNA - RNA - protein, trong khi RNA virus (retrovirus) chuyển mã ngược trước tiên RNA của nó sang DNA, rồi mới tiếp tục mô thức central dogma.

Bảng 2-1 : Mã di truyền

5'	Ký tự số hai				Ký tự số ba
	U	C	A	G	3'
Ký tự số một	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G
	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC His	CGC Arg	C
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A
	AUG Met	ACG Thr	AAG Lys	AGG Arg	G
	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
	GUG ^b Val	GCG Ala	GAG Glu	GGG Gly	G

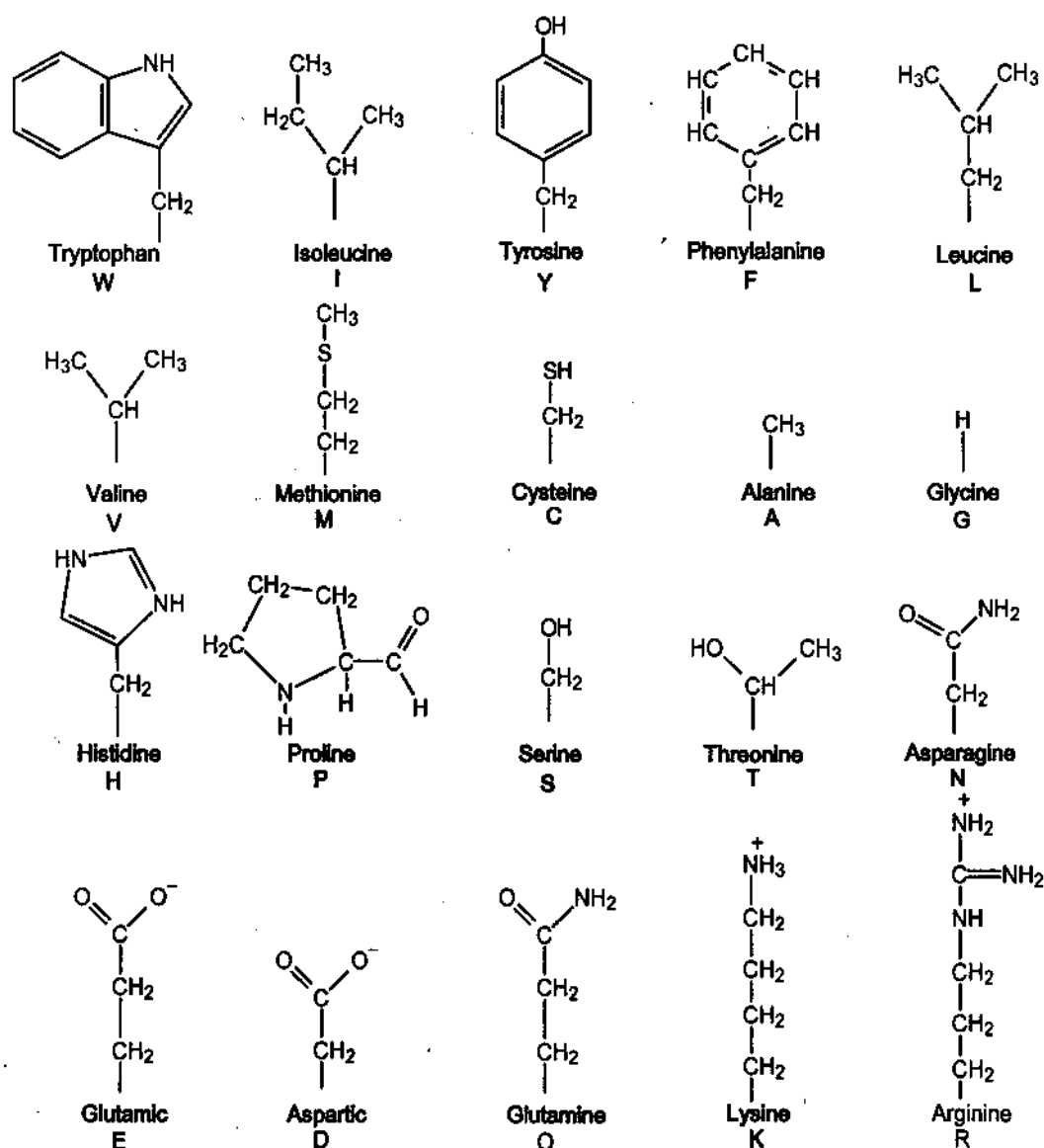
Ghi chú : Mã di truyền UGA (stop) cho Trp trong ty thể bộ của động vật có xương sống, UAA (stop) cho Gln trong protozoa có tiêm mao, CUN (leu) cho Thr trong ty thể bộ của men yeast, và AUA (Ile) cho Met trong ty thể bộ của động vật có xương sống, UGA (stop) cho selenocysteine (SeCys) trong vi khuẩn và một vài động vật có vú.

^b Mã di truyền cho fMet nếu nó ở vị trí bắt đầu.

2.2. CẤU TRÚC ENZYME

2.2.1. Cấu trúc sơ cấp

Sequence của amino acid cung cấp cho chúng ta biết thông tin cơ bản về bản chất của protein, như cấu trúc, cơ chế hoạt động, nguồn tiến hóa, và sự phân loại theo nhóm họ enzyme. Những thông tin sơ khởi này cũng giúp cho chúng ta phát hiện gen cần nghiên cứu, nhân bản vô tính ở mức độ phân tử (cloning), và những kỹ thuật gen khác trong kỹ thuật recombinant DNA.



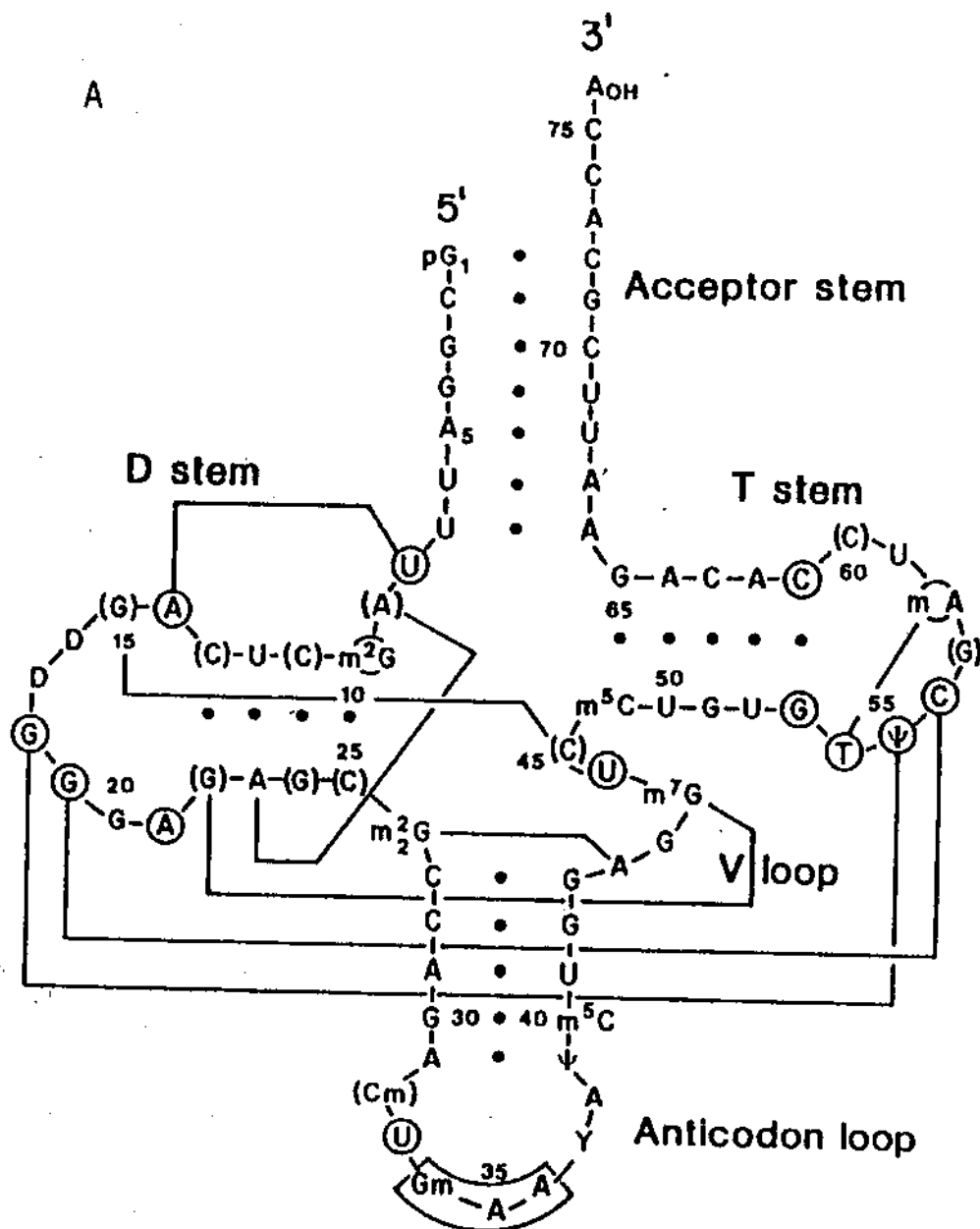
Hình 2.1 : Cấu trúc của amino acid và chữ viết tắt của nó. Xếp theo mức độ ưa nước theo thứ tự; các amino acid kỵ nước (hydrophobic) ở trên và các amino acid (hydrophilic) ở dưới

Bảng 2-2 : Danh sách của 20 amino acid

Amino acid	Ký hiệu và MW	pK _a			
		- COOH ⁺	-NH ₂	R trong	
				aa	protein
Alanine	Ala, 89	2.3	9.8	-	
Arginine	Arg, 174	1.8	9.0	12.5	≥1.2
Asparagine	Asn, 132	2.0	8.8	-	
Aspartic acid	Asp, 133	2.0	9.9	3.9	4.4-4.6
Cysteine	Cys, 121	1.8	10.8	8.3	8.5-8.8
Glutamic acid	Glu, 147	2.2	9.6	4.2	4.4-4.6
Glutamine	Gln, 146	2.2	9.1	-	
Glycine	Gly, 75	2.3	9.7	-	
Histidine	His, 155	1.8	9.2	6.0	6.5-7.0
Isoleucine	Ile, 131	2.3	9.7	-	
Leucine	Leu, 131	2.3	9.6	-	
Lysine	Lys, 146	2.2	9.2	10.8	10.0-10.2
Methionine	Met, 149	2.2	9.2	-	
Phenylalanine	Phe, 165	1.8	9.1	-	
Proline	Pro, 115	2.0	10.6	-	
Serine	Ser, 105	2.2	9.2	-	
Threonine	Thr, 119	2.6	10.4	-	
Tryptophan	Trp, 204	2.4	9.4	-	
Tyrosine	Tyr, 181	2.2	9.1	10.1	9.6-10.0
Valine	Val, 117	2.3	9.6	-	

Chuỗi mã số cấp của protein được xác định bằng hàng loạt các qui trình tiêu chuẩn như sau: (i) phân cắt protein bằng enzyme hay bằng hóa học thành nhiều peptide, (ii) phân biệt những peptide tự do hay được phóng thích (derivatized), (iii) giảm số gốc kể cả đầu N và đầu C (sequential degradation), (iv) phân lập amino acid bằng sắc ký khí và quang phổ hấp thụ.

2.2.2. Cấu trúc thứ cấp (bậc hai)

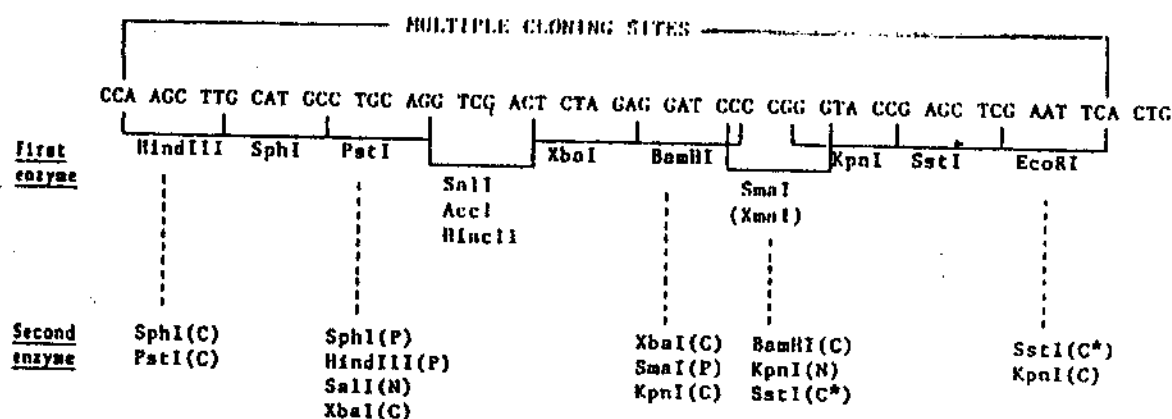


Hình 2.2 : Cấu trúc của tRNA. Cấu trúc bậc hai và cấu trúc 3-D tRNA (Phe) của men. (A) Các base hình thành với cầu nối H theo kiểu bậc ba, dính với nhau theo đường thẳng. Vòng tròn và ngoặc đơn biểu thị base được bảo quản, và bảo quản một nửa, theo thứ tự. (B) cấu trúc 3-D, góc đường phosphatate có dạng một ống nghiêng liên tục. Nối base là một gạch dài, những base đơn độc là gạch ngắn. Cầu nối H theo kiểu bậc ba cho thấy gạch nối giữa các base được phủ kín.

Cấu trúc thứ cấp là hình dạng của dây polypeptide dưới trạng thái ổn định nhất, thường là dạng xoắn ốc. Chuỗi polypeptide của một protein nào đó giả định có những cấu trúc có tính chất thứ tự theo kiểu phân biệt (distinctive ordered), được gọi là cấu trúc thứ cấp (secondary structure). Kiểu xoắn α -helix (theo bên phải hoặc trái) là một cấu trúc có thứ tự, trong đó trục xương sống của nhóm amide (NH và CO) tạo thành một cầu H, với một nhóm amide thứ ba nằm bên ngoài nó. Có 3,6 residues trong mỗi lần xoay qua bên phải (3,6 - fold), tạo một pitch vào khoảng 5,4 amstrong. Đường kính của helix trục xương sống khoảng 6 amstrong.

Một dạng khác của cấu trúc thứ tự là " β sheet" (tấm beta). Trong cấu trúc này, một chuỗi phát triển tạo thành các dây nối với nhau giữa các gốc (interstrand) thay vì nối bên trong gốc (intrastrand) thông qua cầu nối hydrogen, với một hoặc hai dây bắt song song hoặc không song song.

Thêm vào đó là một dạng cấu trúc thứ cấp theo kiểu quán ngẫu nhiên (random coil), được gọi là cấu trúc không cân bằng (nonregular structure). Random coil bao gồm cả việc quán (turn) và tạo vòng (loop), đóng góp một vai trò vô cùng quan trọng trong chức năng protein. Hình 2-2 là một ví dụ của cấu trúc bậc hai và 3D của tRNA.



Hình 2.3 : Phản ứng tiêu hóa gấp đôi (double digestion) tại những vị trí cloning của plasmid pUC19

2.2.3. Cấu trúc tam cấp

Cấu trúc tam cấp là dạng dây xoắn polypeptide gấp khúc lại trong không gian để tạo thành đại phân tử protein. Một protein muốn có hoạt tính sinh học cao, phải có cấu trúc rất chuyên biệt theo kiểu 3-D hay còn gọi là "tertiary" (tam cấp). Hình thành một cấu trúc theo không gian ba chiều 3-D yêu cầu nhiều cải tiến có tính chất hậu giải mã (post-translational).

Enzyme có tính chất single polypeptide của một phân tử có kích thước (≥ 20 kDa) thường được tạo thành từ hai hoặc nhiều domain có tính cấu trúc. Thuật ngữ "domain" liên

quan đến ý nghĩa đơn vị có hình dáng tự gấp lại, compact, với một thể tích lớn trong một bề mặt cực nhỏ.

Cấu trúc 3-D của enzyme hay protein, vừa đơn độc, vừa phức hợp với các chất cơ bản, chất ức chế, hoặc cofactor, được xác định bằng X-quang trên tinh thể.

2.2.4. Cấu trúc tứ cấp

Cấu trúc thứ cấp là cấu trúc gồm nhiều dây polypeptide tam cấp giống nhau hay khác nhau, cùng với nhân hoạt động không phải là protein, tạo nên phân tử protein hoàn chỉnh. Khi có hai hoặc nhiều đơn vị cấu trúc tam cấp (còn được gọi là subunit) tương tác với nhau, hình thành một thể không có tính đồng hóa trị (noncovalent), cấu trúc như vậy được gọi là "quaternary" (tứ cấp). Cách thức chúng tương tác với nhau giữa các subunit, và đặc tính của những subunit này hết sức đa dạng.

Một sự thay đổi nào đó trong cấu trúc tứ cấp đều có thể hiểu là những subunit đang di động đến một vị trí tương ứng khác. Mỗi subunit có thể làm hoàn thiện khả năng của cái khác, nhờ đó thúc đẩy mức độ xúc tác bên trong. Những enzyme có tính chất nhiều subunit như vậy sẽ giúp nó thực hiện chức năng thích ứng với yêu cầu cân bằng trong quá trình biến dưỡng.

2.3. ẢNH HƯỞNG CỦA CẤU TRÚC ĐẾN CHỨC NĂNG ENZYME

2.3.1. Vị trí hoạt động

Chức năng xúc tác của một enzyme thường được xem xét trong cấu trúc tam cấp của nó, nghĩa là cấu trúc 3-D (không gian ba chiều). Chức năng xúc tác của một enzyme được xác định bởi một ít nhóm chức năng đặc biệt của amino acid, nhóm này nằm tại vị trí được gọi là vị trí hoạt động (active site) hay trung tâm hoạt động (active center). Cấu trúc ba chiều, hình dạng, và bản chất của nhóm chức năng hợp thành vị trí hoạt động, theo nguyên tắc bổ sung lẫn nhau, hoạt chất (substance) có thể được dùng làm cơ chất (substrate), còn gọi là sự chuyên biệt về cơ chất. Điều có thể xem quan trọng hơn là: phản ứng hóa học xảy ra theo kiểu nào? Phản ứng có tính đặc thù hay không?

Các residues ở vị trí hoạt động làm việc như một tập hợp nhằm tạo ra một môi trường đồng nhất (microenvironment) mà chúng ta sẽ không tìm thấy bất cứ đâu trên phân tử enzyme. Nhóm có chức năng đặc biệt (thí dụ gốc OH của Ser trong xúc tác serine protease) có thể có phản ứng bất thường do sự rối loạn trong các hợp chất hóa học. Gốc ϵ -NH₂ của Lys ở vị trí hoạt động của aspartate aminotransferase tham dự vào tiến trình vận chuyển proton, nhờ hoạt động như một base có tính nội sinh (endogenous), với một nồng độ hữu hiệu 10⁻⁶ M ethylamine. Mặc dù vì môi trường có tính đồng nhất, nhưng một vị trí hoạt động có thể được hình thành nhờ sự kết hợp của nhiều amino acid trong những enzyme khác nhau. Vị trí hoạt động thường tạo ra một môi trường hữu cực đặc biệt (polar) nhiều

hơn trường hợp môi trường vô cực (nonpolar). Vị trí hoạt động thúc đẩy sự ổn định có tính chất tĩnh điện của trạng thái vận chuyển ion bằng hình thức sử dụng dung môi mạnh hơn nhiều lần sử dụng nước (Eun 1996).

Các enzyme như DNA polymerase cần có rất nhiều cơ chất gắn vào. Trong những enzyme có tính chất multisubstrate, thì mỗi substrate thường chiếm một subsite nằm trong trung tâm hoạt động. Các gốc hóa học trong một vị trí hoạt động nào đó đều rất quan trọng trong tiến trình xúc tác, có gốc hóa học làm nhiệm vụ gắn các cơ chất lại, còn phần lớn đều làm nhiệm vụ xúc tác hóa học. Nhóm cơ chất của enzyme có tính chất không đồng hóa trị (noncovalent) được gọi là Michaelis-Menten complex.

2.3.2. Sự chuyên tính của cơ chất

2.3.2.1. Cơ chất và sự tương đồng

Một enzyme có thể gắn rất chặt với một cơ chất này và không chặt với một cơ chất khác. Lực gắn còn gọi là "binding affinity" có thể được so sánh nhờ phóng đại các thể ổn định gắn vào (binding constant), còn gọi là sự kết hợp. Thể ổn định này là bản sao chụp ngược của thể dissociation constant (không kết hợp), nó là thông số đo lường về động thái học của enzyme.

Những hợp chất có cấu trúc giống như cơ chất thật, nhưng không thể tác động xoay chiều khi có enzyme, được gọi là cơ chất giả (pseudosubstrate), chất tương đồng (substrate analogs), hay chất ức chế (inhibitors). Mức độ phóng đại tương đối của lực gắn đối với hàng loạt cơ chất và/ hay analog thường phản ánh chiều vật lý trong không gian, và bản chất lý hóa của vị trí hoạt động và tương tác giữa enzyme-cơ chất. Hầu hết các analogs của cơ chất gắn với enzyme không chặt lắm, nhưng có vài analog như transition-state analogs gắn với enzyme chặt hơn cơ chất rất nhiều, kết quả tạo ra lực cản đặc biệt đối với enzyme. Thí dụ thứ hai vanadyl ribonucleoside complex (VRC) là một chất cản transition-state đối với RNase A.

2.3.2.2. Sự chuyên tính lập thể (stereospecificity)

Trong một vài trường hợp ngoại lệ, enzyme có chức năng chuyên tính về lập thể, đối với việc gắn vào cơ chất. Nhờ tính chất này, enzyme phân biệt được các cơ chất có tên gọi là những isomer quang học (optical isomers, enantiomers, stereoisomers). Trong tất cả các amino acid, ngoại trừ Gly có chuỗi nối là nguyên tử hydrogen, thì nguyên tử α carbon có tính chất không đối xứng và nguyên tử này có độ triển quang về ph. trái.

2.3.3. Các họ của enzyme

2.3.3.1. Xếp họ theo cấu trúc hoặc chức năng

Xét theo mức độ xúc tác (catalytic turnover rate), tiến trình xúc tác (pathway), người ta phân ra các họ enzyme. Những thành viên trong một họ có một kiểu sequence hoặc một

nhóm chức năng khác với các họ khác.

So sánh những sequence của amino acid còn được gọi là công việc tìm kiếm tương đồng (homology search). Công việc này thường cung cấp những thông tin đầu tiên cho phép việc phân loại enzyme thành các họ. Nó cũng cho phép dự đoán chức năng của một enzyme mới. Một họ enzyme thường có ít nhất một vùng hoạt động. Thí dụ điển hình trên họ serine protease mà những thành viên của nó là trypsin, chymotrypsin, elastase, subtilisin - thường có 3 residues ở vị trí hoạt động, Asp/His/Ser, người ta gọi đó là "catalytic triad" hay "charge relay system". Thí dụ khác về đặc điểm bảo tồn của lĩnh vực xúc tác đối với họ protein kinase rất đa dạng. Trong họ kinase, vị trí gắn ATP (hay phosphate-anchoring) đại diện cho một kiểu sequence của Gly-X-Gly-X-X-Gly.

2.3.3.2. Isoenzyme

Isoenzyme hay isozyme là một nhóm enzyme xúc tác cùng phản ứng, nhưng có nhiều dạng enzyme khác nhau và hiệu quả xúc tác khác nhau. Isozyme được phân biệt nhờ chạy điện di trên gel.

Isozyme được dùng làm marker trong việc tìm kiếm gen trong phân tích mức độ đa dạng về di truyền của loài. Nếu những liên kết như vậy, người ta còn dùng thuật ngữ "recombinant enzyme" để chỉ hiện tượng này. Tuy nhiên số lượng isozyme khá ít không đáp ứng nhu cầu lập bản đồ di truyền.

Thuật ngữ isozyme mô tả các dạng phân tử khác nhau của những enzyme với cùng đặc tính về cơ chất. Nó cho biết bản thân của những alen khác nhau ở một locus biết trước, hoặc bản thân của những loci khác nhau. Với sự phát triển của kỹ thuật điện di, thí dụ nghiên cứu sự ổn định nhiệt học của isozyme, người ta có thể chẩn đoán được một phần nào sự biến dị ở mức độ DNA.

Các isozyme thường được sử dụng trong phân tích di truyền huyết thống là:

- Acid phosphatase
- Alanine aminopeptidase
- Alcohol dehydrogenase
- Arginine aminopeptidase
- Catalase
- Esterase
- Glutamate oxalo acetate transamiase
- Isocitrate dehydrogenase
- Leucine aminopeptidase

- Malic enzyme
- Peroxidase
- Phosphogluconate dehydrogenase
- Phosphoglucose isomerase
- Shikimate dehydrogenase

2.3.4. Phân loại enzyme và định danh

Hiện nay, việc phân loại và định danh enzyme dựa theo tiêu chuẩn của IUBMB viết tắt từ chữ International Union of Biochemistry and Molecular Biology, trên cơ sở phản ứng chuyên tính của enzyme. Mỗi enzyme khi được phân loại được kèm theo một số mã hiệu khởi đầu bằng chữ EC (Enzyme Commission). Số thứ nhất cho biết một trong sáu lớp (class) của enzyme. Chữ số thứ hai, ba và tư cho biết subclass, sub-subclass, và số serie của enzyme trong sub-subclass, theo thứ tự.

Có 6 lớp của enzyme :

1. Oxidoreductase như dehydrogenase, oxidase xúc tác trong phản ứng oxid hóa khử
2. Transferase trong phản ứng vận chuyển các donor và acceptor
3. Hydrolase xúc tác trong phản ứng thủy phân nối C-O, C-N, C-C và các nối khác
4. Lyase trong phản ứng phân cắt C-C, C-O, C-N và các cầu nối khác bằng cách loại trừ (elimination), thêm nhóm để có cầu nối đôi
5. Isomerase xúc tác trong phản ứng thay đổi cấu trúc với một phân tử
6. Ligase xúc tác trong phản ứng gắn lại hai phân tử nhờ thủy phân cầu pyrophosphate trong ATP, hoặc một triphosphate tương tự

2.3.4. Thành phần không phải protein của enzyme

- Cofactor: là một phần trong vị trí hoạt động, thí dụ như pyridoxal phosphate là cofactor của nhóm enzyme làm nhiệm vụ vận tải (aminotransferase)
- Coenzyme: có tính chất đồng xúc tác, chất này chuyển đổi thành sản phẩm mới kèm theo sau mỗi turnover của cơ chất. Thí dụ coenzym là pyridine nucleotide (NAD^+ và NADP^+), flavin (FMN, FAD), tham dự vào nhiều phản ứng oxid hóa - khử.
- Nhóm prosthetic: trái với coenzyme và cofactor làm chức năng chuyển đổi hóa học trong tiến trình xúc tác, nhóm prosthetic là một hợp chất ổn định, không thay đổi. Nó có xu hướng gắn với enzyme đồng hóa trị hoặc không đồng hóa

trị, mà lực gắn này thường chặt hơn lực gắn của cofactor và coenzyme. Thí dụ như những ion kim loại Zn^{++} trong dehydrogenase và alkaline phosphatase, Fe^{++} trong oxidoreductase. Thí dụ như hợp chất hữu cơ TTQ (tryptophan tryptophylquinone) trong methylamine dehydrogenase (EC1.4.99.3)

2.4. RESTRICTION ENDONUCLEASE

Trong phân tích genome, enzyme làm nhiệm vụ cực kỳ quan trọng là nhóm enzyme, với thuật ngữ chung "restriction endonuclease". Đó là nhóm của DNase, ghi nhận các sequence đặc biệt của nucleotide, và cắt DNA tại vị trí rất đặc biệt. Người ta xem nó như chìa khóa để nghiên cứu kỹ thuật recombinant DNA.

Nó được nghiên cứu lần đầu vào những năm 1950. Nó bao gồm một phần của tính chất giới hạn (restriction) viết tắt là R, và một phần của tính chất cải tiến (modification) viết tắt là M. Hệ thống R-M trong vi khuẩn giúp nó bảo vệ chống lại sự xâm nhập của thực khuẩn thể, và những nhân tố lạ về di truyền. Từ kiểu hình được phân lập đầu tiên, hệ thống R-M bây giờ có thể sẵn sàng cho những phân tích di truyền rất có hiệu quả. Hệ thống R-M của vi khuẩn phá hủy những phân tử DNA lạ xâm nhập vào tế bào bằng con đường lây nhiễm, tiếp hợp, chuyển nạp. Hệ thống R-M của vi khuẩn có xu hướng tạo ra một thể cân bằng ở sinh vật nhân giả (prokaryotic) như là một hệ thống miễn dịch.

Hệ thống R-M được tìm thấy trong vi sinh vật, chủ yếu ở vi khuẩn. Nó rất đa dạng ngay cả trong cùng một tế bào. Số restriction enzyme được tư liệu hóa cho đến nay trên 2100, trong đó có 17 thuộc type I, 179 type II, 4 type III rất chuyên tính, bên cạnh đó hơn 190 nhóm DNA modification methyltransferase đã và đang được định tính (Kita và ctv. 1989).

Hệ thống R-M thường có hai hoạt động chính: (a) một restriction endonuclease viết tắt là R-ENase rất chuyên tính tại một vị trí nào đó, nó có nhiệm vụ tiêu hóa DNA ngoại sinh (exogenous DNA), (b) một modification methylase của DNA, còn được gọi là methyltransferase, viết tắt M-MTase, có sự chuyên tính về chuỗi mã rất đồng nhất. M-MTase có nhiệm vụ cải tiến và bảo vệ DNA nội sinh (endogenous DNA) không bị tiêu hóa bởi R-ENase.

2.4.1. Các dạng (type) của hệ thống R-M

Có 4 dạng : type I, type II, type III và type IV dựa trên cơ sở hợp phần của enzyme, nhu cầu về cofactor, tính đối xứng của sequence, và đặc tính phân cắt (cleavage) (Bảng 2-3). Tất cả hoạt động của MTase cần có AdoMet như là nhóm methyl của cơ chất cho.

Bảng 2-3 : Phân loại hệ thống R-M

Đặc điểm	Type I	Type II		Type III	Type IV
		Prototype	Type IIS		
cấu trúc hoạt động R-M	enzyme đơn 3 s.u. complex	phân loại enzyme		enzyme đơn 2 s.u. complex	
		R: dimer M: monomer	R: monomer M: monomer		
Cofactor	Mg ²⁺ ATP	Mg ²⁺		Mg ²⁺ ATP	Mg ²⁺ (AdoMet)
Vị trí xác định / phân cắt	Không đối xứng, khoảng cách đa dạng	Palindromic cùng vị trí	Không đối xứng khoảng cách xác định 3'	Không đối xứng 25-27 bp đến 3'	Không đối xứng 14bp đến 3'
Methyl hoá	Hai dây	Hai dây	Một MTase trên mỗi dây	chỉ một dây	Hai dây

- Type I: R-M enzyme là multienzyme, có 3 subunit không đồng nhất: R, M, và S. Subunit S xác định sự chuyên tính đối với R và M. Type I R-M cần Mg²⁺ và ATP như là cofactor, và cắt DNA chưa được cải tiến tại vị trí tương đối ngẫu nhiên. Enzyme của type I thủy phân ATP trong hoạt động restriction, và methyl hóa rất chuyên tính DNA của nó
- Type II: hệ thống R-M gồm có hoạt động của cả hai R-ENase và M-MTase. R-ENase thường là một dimer, với hai subunit đồng nhất, trong khi M-MTase là những enzyme monomeric. R-ENase của type II chỉ cần có Mg²⁺ giúp cho hoạt động tiêu hóa của nó. R-ENase và M-MTase của type II ghi nhận cùng sequence rất chuyên tính của nucleotide. Nhóm enzyme này cắt DNA tại sequence như vậy, trong khi một nhóm nhỏ của enzyme type IIS, cắt DNA ở khoảng cách với vị trí target một độ dài đã được xác định.
- Type III: có 2 subunit không đồng nhất. Hoạt động tiêu hóa cần Mg²⁺ và ATP. Enzymes của loại này ghi nhận chuỗi mã ngắn, không có tính chất palindromic (5'-3'), và cắt DNA tại vị trí cố định của nó. Enzyme của loại hình này không thủy phân được ATP. Hoạt động tiêu hóa không cần AdoMet, nhưng nó cần để kích hoạt ENase.

- Type IV: đại diện cho loại hình trung gian trong con đường tiến hóa giữa enzyme của type II và type III. Thí dụ như *Eco* 571 có R-ENase và M-MTase phân biệt nhau, nhưng monomeric R-ENase có chứa hoạt động methylase bổ sung. Hoạt động restriction hết hợp với methylase chưa đủ mạnh để bảo vệ DNA của tế bào chủ *in vivo*. Hoạt động restriction cần được kích hoạt nhờ AdoMet.

2.4.2. Định danh

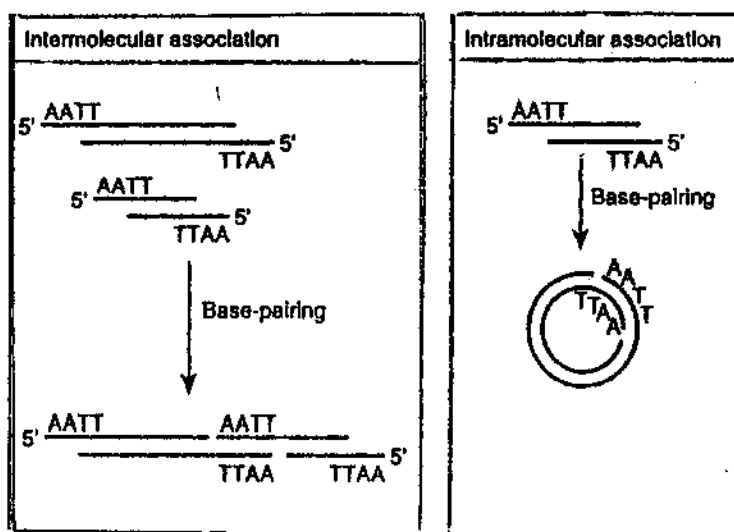
Bảng 2-4 : Một vài restriction endonuclease và vị trí phân cắt của nó

Nguồn vi khuẩn	Ký hiệu enzyme	Sequence	Ghi chú
		5'→3' 3'←5'	
<i>Haemophilus aegyptius</i>	<i>HaeIII</i>	GG CC CC GG	1
<i>Staphylococcus aureus</i> 3A	<i>Sau3AI</i>	GATC CTAG	2
<i>Bacillus amyloliquefaciens</i> H	<i>BamHI</i>	G GATC C C CTAG G	2
<i>Escherichia coli</i> RY13	<i>EcoRI</i>	G AATT C CTTAA G	2, 6, 7
<i>Haemophilus influenzae</i> Rd	<i>HindII</i>	GTPy PuAC CTPu PyTG	1,5
	<i>HindIII</i>	A AGCT T T TCGA A	2
<i>Providencia stuartii</i>	<i>PstI</i>	CTGCA G G ACGT C	3
<i>Serratia marcescens</i>	<i>SmaI</i>	CCC GGG GGG CCC	1
<i>Xanthomonas malvacearum</i>	<i>XmaI</i>	C CCGG G G GGCC C	2
<i>Moraxella bovis</i>	<i>MboII</i>	GAAGAN ₈ CTTCTN ₇	4

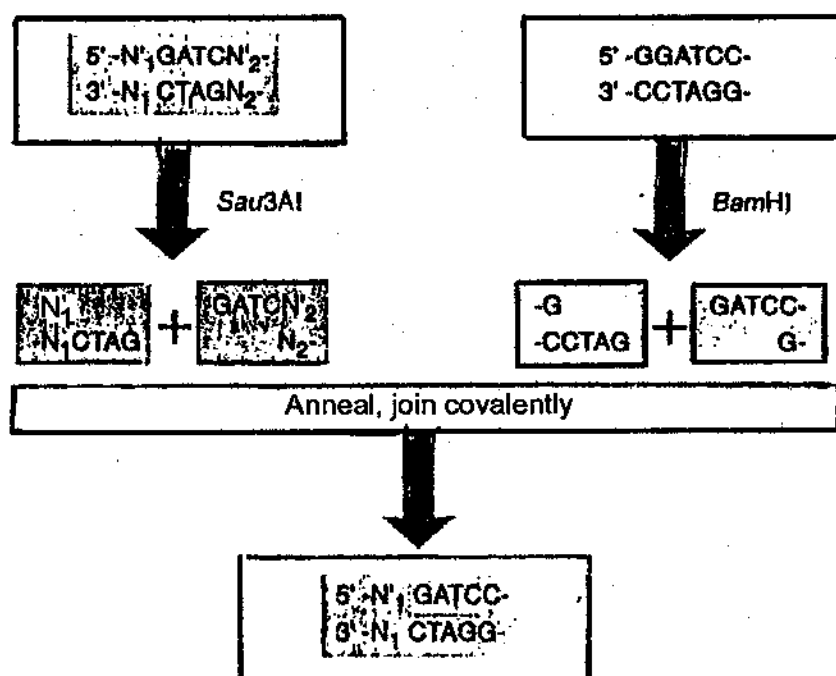
Smith và Nathan (1973) đã đề nghị một hệ thống định danh thống nhất như sau:

- Tên loài của sinh vật làm ký chủ được xác định, xong lấy chữ cái đầu tiên của genus cộng thêm hai chữ cái đầu tiên của tên loài = 3 chữ, viết theo kiểu in nghiêng (italic). Thí dụ *Escherichia coli* = *Eco* và *Haemophilus influenzae* = *Hin*.
- Nòi của vi khuẩn được viết bằng chữ in hoa, sệt xuống hàng dưới (subscript). Thí dụ *Eco_K*. Trong trường hợp hệ thống R-M được chuyên tính bởi virus hay plasmid, người ta dùng chữ viết tắt của tên loài ký chủ và nguyên tố ngoại nhiễm (extrachromosomal) được xác định theo sau chữ ấy, viết xuống hàng dưới. Thí dụ *Eco_{PI}*, *Eco_{RI}*.
- Khi một dòng ký chủ nào đó có nhiều hệ thống R-M khác nhau, người ta dùng số la mã đánh dấu ở phía sau. Thí dụ *Hin_I*, *Hin_{II}*, *Hin_{III}*, v.v.. Số la mã này không phải là biểu thị loại hình enzyme như đã nói trên (type I, type II,...)
- Tất cả mọi restriction enzyme đều có tên chung là "endonuclease R", nhưng bên cạnh đó, nó còn mang thêm tên hệ thống, như *R.Hin_I*. Tương tự, modification enzyme có tên chung là "methylase M" theo sau đó là tên hệ thống. Modification enzyme của *H. influenzae* Rd tương hợp với endonuclease *R.Hin_I* sẽ được ghi là *M. Hin_I*.

Trong thực tế, người ta muốn việc sử dụng khi viết sao cho đơn giản, (a) chữ viết nhảy xuống dưới không cần thiết, toàn bộ chữ viết tắt được ghi cùng một hàng, thí dụ *HindIII*, và (b) khi biết rõ là chỉ có restriction enzyme, ký hiệu tượng trưng cho endonuclease R bị xóa bỏ. Xem bảng 2-4 về ký hiệu viết tắt của restriction endonuclease enzyme (Old và Primrose 1995).



Hình 2.4 : Đầu dính theo kiểu so le (sticky hoặc cohesive) của đoạn phân tử DNA tạo ra nhờ tiêu hóa với *EcoRI*



Hình 2.5 : Tạo ra một vị trí lai (hybrid site) bằng cách gắn vào đầu sticky, nhờ *Sau3AI* và *BamHI*

1. Cắt theo kiểu blunt end (dạng thẳng hai đầu so với dây đôi)
2. Cắt theo kiểu sticky end hay cohesive end, đoạn cuối dây đôi so le như một ngàm gỗ, tại vị trí 5' (hình 2-4, 2-9)
3. Cắt theo kiểu sticky end, tại vị trí 3'
4. Đây là type IIs của enzyme. Nó không cắt trong vùng mã ghi nhận (recognition sequence), mà cắt sequence nào đó ở bên mặt N là nucleotide
5. Pu là purine (A hoặc G), Py là pyrimidine (C hoặc T). Mọi cấu trúc có cặp base như vậy, đều bị cắt.
6. Trong một điều kiện (lực ion yếu, pH kiềm, hoặc 50% glycerol), sự chuyên tính của *EcoRI* giảm. Đặc tính này được gọi là hoạt động của *EcoRI*-star. Hoạt động star bị ngăn cản bởi *p*-chloromercuribenzoate, trong khi hoạt động bình thường không mẫn cảm.
7. *EcoRI* và *EcoRII* cắt $\downarrow CC(A/T)GG$, được chế biến từ hai dòng vi khuẩn mang tính chất kháng sinh khác nhau, mang những factor (plasmid) có mã di truyền của restriction enzyme. Hai factor chứa tính kháng này là RI và RII

2.4.3. Chuỗi mã xác định (Recognition Sequence)

Đây là thuật ngữ thường được áp dụng trong kỹ thuật recombinant DNA. Chữ recognition sequence ở đây có nghĩa là enzyme có tính chất chuyên biệt với một sequence

nào đó, được xác định bởi chính nó, và nó sẽ cắt nucleotide đúng vị trí xác định mà nó tìm thấy, vị trí này còn được gọi là "target site", với số nucleotide trong chuỗi được xác định.. Từ đó, DNA của genome sẽ được cắt ra thành những đoạn ngắn, hoặc dài, có độ lớn bp khác nhau. Đó chính là chìa khóa trong công nghệ di truyền, nhằm sáng tạo ra những marker, áp dụng trong phân tích genome sau này.

Type II của restriction endonuclease được người ta chú ý nhiều nhất. Nó có recognition sequence theo kiểu tetra-, penta-, và hexanucleotide. Những sequence như vậy được viết từ trái sang phải theo hướng 5'-3' (palindromic). Thí dụ 5'-GAATTC-3'. Chuỗi mã xác định có trục đối xứng có tính chất vòng kín (rotational). Thí dụ recognition sequence của *Eco* RI là :

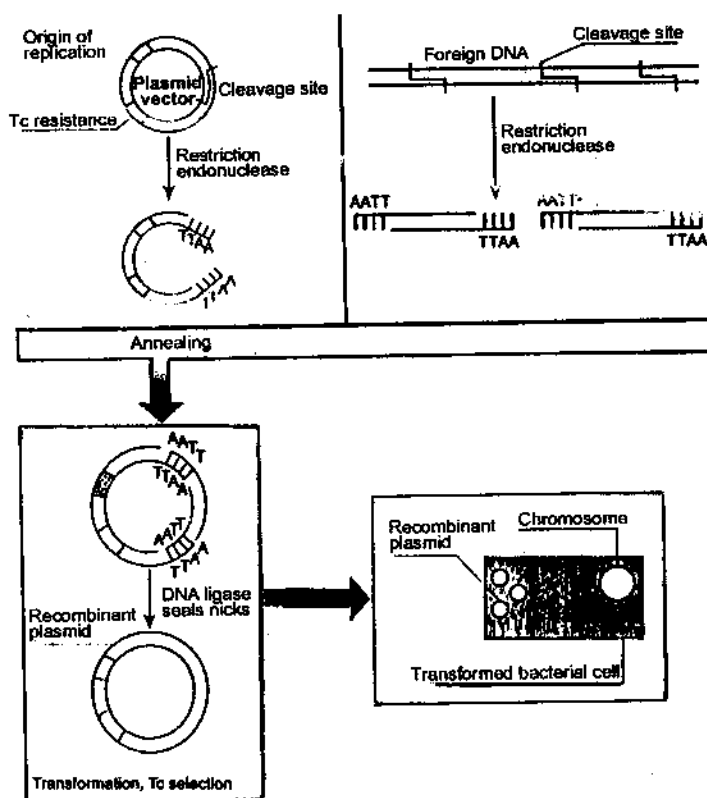
5'-GAATTC-3'

3'-CTTAAG-5'

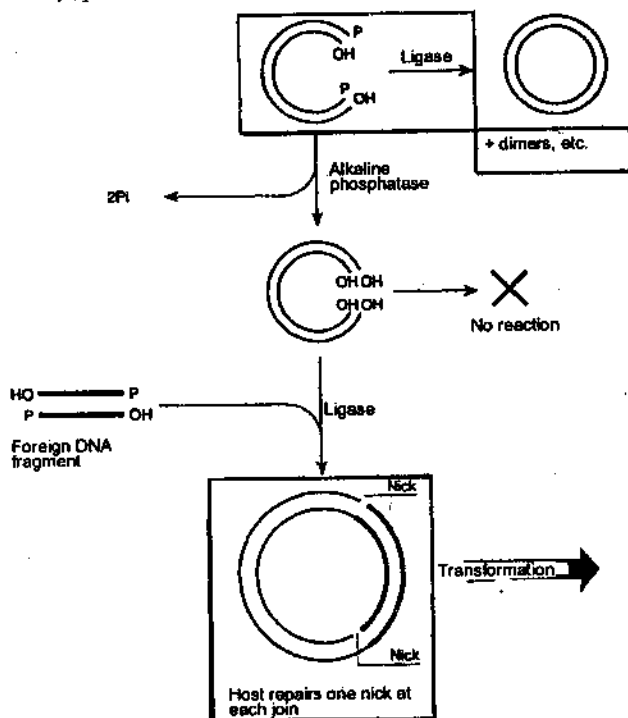
Về lý thuyết, có 16 tetranucleotide đối xứng và 64 hexanucleotide. Tuy nhiên chỉ có 50% thực sự hoạt động tại các target site của nó, đối với những restriction endonuclease được biết .

Bảng 2-5: Liệt kê các restriction endonuclease xếp thứ tự theo bản chất vị trí cắt của nó

5'- end		3'-end		Blunt end	
Enzyme	Sequence	Enzyme	Sequence	Enzyme	Sequence
<i>Taq</i> I	T/CGA	<i>Pst</i> I	CTGCA/G	<i>Alu</i> I	AG/CT
<i>Cl</i> aI	AT/CGAT	<i>Sac</i> I	GAGCT/C	<i>Fnu</i> DII	CG/CG
<i>Mbo</i> I	/GATC	<i>Sph</i> I	GCATG/C	<i>Dpn</i> I	GA/TC
<i>Bgl</i> II	A/GATCT	<i>Bde</i> I	GGCGC/C	<i>Hae</i> III	GG/CC
<i>Bam</i> HI	G/GATCC	<i>Apa</i> I	GGGCC/C	<i>Pvu</i> II	CAG/CT
<i>Bcl</i> I	T/GATCA	<i>Kpn</i> I	GGTAC/C	<i>Sma</i> I	CCC/GGG
<i>Hind</i> III	A/AGCTT			<i>Nae</i> I	GCC/GGC
<i>Nco</i> I	C/CATG			<i>Hpa</i> I	GTT/AAC
<i>Xma</i> I	C/CCGG			<i>Nru</i> I	TCG/CGA
<i>Xho</i> I	C/TCGAG			<i>Bal</i> I	TGG/CCA
<i>Eco</i> RI	G/AATTC			<i>Mst</i> I	TGC/GCA
<i>Sal</i> I	G/TCGAC			<i>Aha</i> III	TTT/AAA
<i>Xba</i> I	T/CTAGA			<i>Eco</i> RV	GAT/ATC



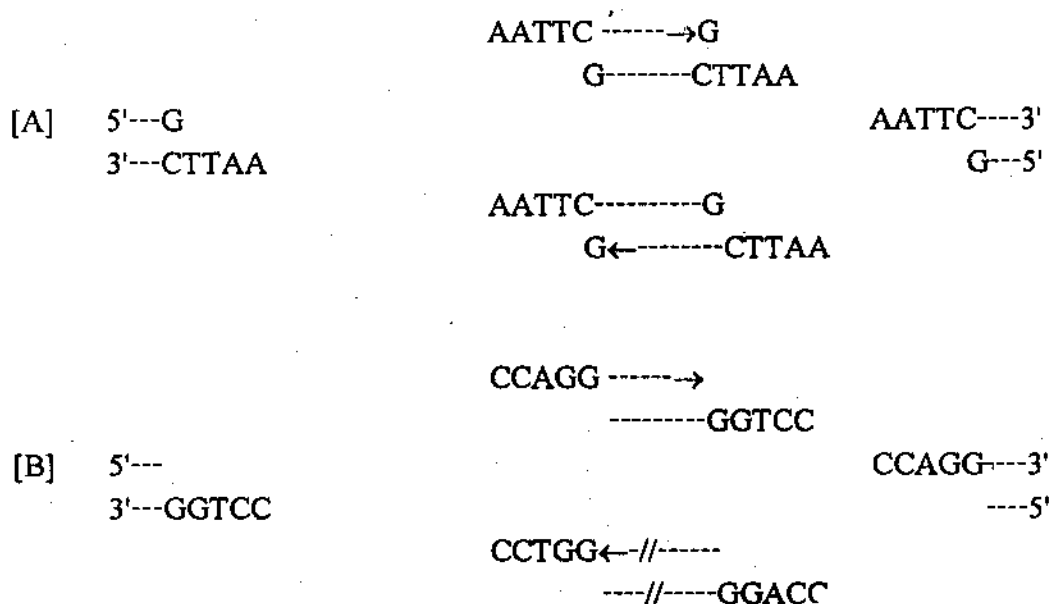
Hình 2-6: Sử dụng DNA ligase để tạo ra một thể recombinant thông qua sự kiện kết hợp các đầu dính, phát sinh do *EcoRI*



Hình 2-7: Ứng dụng alkaline phosphate để ngăn ngừa hiện tượng tái lập lại chu kỳ của plasmid chưa mang DNA lạ

Thủy phân để cắt hai dây đơn DNA bằng type II sẽ xảy ra trong chuỗi xác định của nó. Tiến trình này làm dây đơn ngắn lại. Bảng 2-5 cho thấy danh sách của các restriction enzyme được xếp theo nhóm tùy thuộc vào loại hình termini (kết thúc) vốn là bản chất của nó. Việc hiểu biết cặn kẽ kiểu termini như vậy, có thể sẽ rất có ích cho việc thiết lập những phân tử recombinant DNA mới từ các đoạn DNA nào đó.

Cách sắp xếp đối xứng của các vị trí mà enzyme đã cắt, sẽ có hai hệ quả quan trọng cần lưu ý: (i) thể hệ của những sản phẩm do sự cắt này tạo ra một termini dây đơn, hoàn toàn có tính đối song (antiparallel), nhưng đồng nhất về hình dáng song song, (ii) các đầu DNA của một phân tử được phát sinh nhờ một enzyme nào đó, với một chuỗi xác định nào đó có thể hình thành những cặp base có tính bổ sung với những phân tử DNA khác, làm cho nó có cùng một dạng ở đầu dây DNA (hình 2-6 và 2-7 cho ví dụ về kết quả cắt, nối tại chuỗi mã xác định).



Hình 2.8: Sự sắp xếp của những đoạn DNA với đầu dây đối xứng, hoặc không đối xứng

[A] Vị trí cắt của *EcoRI*: đoạn DNA có thể được gắn vào cả hai hướng mũi tên

[B] Vị trí cắt của *EcoRV*: đoạn DNA có thể được gắn chỉ theo một hướng, không đối xứng

Quan sát có tính chất quan trọng này được thực hiện do Mertz và Davis (1972), trong khi họ nghiên cứu enzyme *EcoRI* (G/AATTC). Enzyme này xác định bốn sequence tương thích khác nhau. Sự tiêu hóa với enzyme như vậy có thể cho ra những termini không đồng nhất, những termini này không thể tái hợp trở lại một cách ngẫu nhiên. Bảng 2-6 cho thấy danh sách của một vài enzyme có nhiều recognition sequence. Thí dụ *EcoRV*, thể hệ của những đầu dây không đồng nhất do sự cắt không đối xứng xảy ra trong suốt quá trình tái tổ hợp *in vitro*.

Bảng 2-6 : Danh sách những enzyme có nhiều recognition sequence

Enzyme	Recognition sequence					Số vị trí trên phân tử		
						pBR322	M13mp8	SV40
						DNA	DNA	DNA
<i>EcoRI</i>	/CC	A	GG			6	7	17
	/CC	T	GG					
<i>AccI</i>	GT/	AG	AC			2	1	1
	GT/	AT	AC					
	GT/	CG	AC					
	GT/	CT	AC					
<i>AvaI</i>	C/	C	CG	A	G	1	2	0
	C/	C	CG	G	G			
	C/	T	CG	A	G			
	C/	T	CG	G	G			
<i>HaeII</i>	A	GC	GC/	C		11	6	1
	A	GC	GC/	T				
	G	GC	GC/	C				
	G	GC	GC/	T				
<i>HindII</i>	GT	C	/	A	AC	2	1	7
	GT	C	/	G	AC			
	GT	T	/	A	AC			
	GT	T	/	G	AC			

Nhiều enzyme có thuật ngữ là "isoschimers" có những vị trí xác định (recognition) đồng nhất. Thí dụ : *HindIII* và *HsuI* (A/AGCTT) chia điểm cắt giống nhau và chuỗi xác định giống nhau. Trái lại *SmaI* (CCC/GGG) và *XmaI* (C/CCGGG) có chung chuỗi xác định, nhưng khác điểm cắt. Vài enzyme có thể khác nhau về chuỗi xác định, nhưng lại tạo

ra những termini có tính chất chồng lấp (overlapping). Thí dụ, tiêu hóa với *Bam*HI (G/GATCC), *Bgl* II (A/GATCT), *Bcl* I (T/GATCA) và *Sau* 3A (/GATC), cho chúng ta một sticky end có mã 5'-GATC-3'. Tất cả những DNA được tạo ra do sự cắt kiểu như vậy với *Bam*HI và *Bgl* II, chúng có thể tái tổ hợp lại với nhau. Trong trường hợp này, thể recombinant sẽ có chuỗi mã 5'-GGATCT-3' và sẽ kháng lại sự tiêu hóa bởi enzyme, bởi vì các base lân cận trong chuỗi xác định dạng hexanucleotide của *Bam*HI và *Bgl* II sẽ rất khác nhau (G/C và A/T, theo thứ tự) do đó, các phân tử recombinant và bố mẹ (DNA gốc) có thể được cắt bởi enzyme *Dpn*I, *Sau* 3A, và *Mbo*I, nhờ vậy nó rất dễ được phân biệt.

2.4.4. Phát hiện và làm thuần khiết restriction endonuclease

Các đoạn DNA có thể được nhìn thấy dưới tia cực tím (UV), nhờ tính huỳnh quang của DNA đã nhuộm với ethidium bromide. Tia sáng ở độ dài sóng 254nm có độ nhạy cảm cao nhất. Tuy nhiên, người ta cũng thấy rằng, phân tử DNA có thể được phục hồi trên gel, với điều kiện độ dài sóng của tia sáng 366nm.

Sự chuyển động của các đoạn DNA trên agarose và polyacrylamide gel tương ứng với logarithm của trọng lượng phân tử của đoạn DNA này.

Có nhiều phương pháp được khuyến cáo để phục hồi DNA từ những môi trường gel như vậy (Southern 1979). Trong đó có 4 phương pháp được thảo luận sau đây có mức độ hữu hiệu cao nhất:

Phương pháp 1 :

Langridge và ctv. (1980) dựa trên tính chất chuyển đổi của đoạn DNA ở dạng muối hexadecyltrimethylammonium. Đây là một chất dễ tan trong dung môi hữu cơ, như butanol. Người ta có thể phân biệt dễ dàng DNA với agarose, duy trì ở trạng thái lỏng. Các đoạn DNA được chuyển vào thể lỏng này bằng cách tăng dần nồng độ NaCl, bên cạnh đó muối ammonium thể quaternary được loại ra nhờ ly trích chloroform. Điều quan trọng là sử dụng agarose có độ melting thấp. Bởi vì tấm gel mỏng đang chứa các đoạn DNA được điện di, có thể nóng 70°C, như vậy DNA có thể bị ly trích ra khỏi pha lỏng.

Phương pháp 2 :

Các tấm mỏng agarose có chứa DNA được đặt vào đỉnh cột agarose, trong một ống thủy tinh hình trụ, ống này được hàn kín bởi agarose và tiếp nhận một lớp mỏng hydroxyapatite ở trên. DNA này có thể được chuyển vào lớp hydroxyapatite, nhờ dòng điện di tiếp sau đó, trong một buồng điện di cột chuẩn. Lớp hydroxyapatite có DNA này được trải lên từ từ đỉnh của một sephadex G25, hoặc cột biogel P30, được điều chỉnh thường xuyên bằng thể đệm là muối ở nồng độ thấp. DNA được xử lý tiếp tục với sodium phosphate 0,7M, pH = 6,8. Từng đoạn DNA riêng lẻ được hòa tan đầy đủ trong agarose gel. Người ta đưa hydroxyapatite này vào một slot. Phân tử DNA có thể được chuyển trực tiếp vào lớp mỏng hydroxyapatite nhờ điện di liên tục (Tabak và Flavell 1978).

Phương pháp 3 :

Chạy điện di các mẫu DNA trên gel nằm ngang. Các lỗ chứa (giếng) mẫu DNA và chất đệm đặt trên đầu tấm. Điện di liên tục cho đến khi nào các đoạn DNA mong muốn chạy hết trên dây của nó. Với sự có mặt của ethidium bromide, DNA được nhuộm sắc dưới tia cực tím, thành các vệt rất rõ. DNA được hồi phục trong ống thủy tinh đặc biệt Pasteur pipette, trước khi cho nó trở lại thể gel. Trong trường hợp này, một tấm gel mỏng có chứa các đoạn DNA được cắt ra khỏi gel, rồi để nó trong túi dialysis. Túi này được đặt trong một công cụ có tên là "slab gel", theo hướng thẳng góc với dòng điện. Người ta sử dụng tia UV để theo dõi kết quả điện di. Túi dialysis sẽ làm giảm thiểu tối đa sự mất mát DNA do tính chất phân tán trong hộp điện di (McDonnell và ctv. 1977)

Phương pháp 4 :

Các đoạn DNA có kích thước nhỏ hơn 500 bp được trích ra một cách dễ dàng từ agarose gel hoặc polyacrylamide gel bằng cách ủ ấm lâu trong dung dịch muối (500mM ammonium acetate, 10mM magnesium acetate) (Maxam và Gilbert 1980). Vệt của gel matrix có thể được lấy ra bằng phản ứng phenol ngay sau đó

2.4.5. Tính chất đặc biệt của chuỗi mã xác định

Số lượng các enzyme (type II) cần thiết cho phản ứng phân cắt DNA, tùy thuộc vào độ lớn của DNA, hợp phần của những base trong DNA, đặc biệt tỉ lệ GC trong toàn bộ base của DNA. Xác suất của tetranucleotide có 50% AT và 50% GC trong phân tử DNA là $1/4^4$. Tương tự như vậy, xác suất của hexanucleotide là $1/4^6$.

Bảng 2-7: Tần suất của recognition site có tính chất hexanucleotide trong plasmid pBR322 và SV40 DNA

Enzyme	Chuỗi mã xác định	Số chuỗi mã xác định	
		pBR322 DNA	SV40 DNA
		4363bp; 54% GC	5243bp; 40,8% GC
<i>Bam</i> HI	GGATCC	1	1
<i>Eco</i> RI	GAATTC	1	1
<i>Hind</i> III	AAGCTT	1	6
<i>Hpa</i> I	GTTAAC	0	4
<i>Pst</i> I	CTGCA	1	2
<i>Pvu</i> II	CAGCT	1	3
<i>Sma</i> I	CCCGG	0	0
<i>Sac</i> II	CCGCG	0	0
<i>Xma</i> III	CGGCCG	1	0

Thông thường, enzyme có chuỗi mã xác định thuộc dạng hexanucleotide, với đầy đủ 4 base, sẽ có hơn một vị trí xác định (recognition site) trong phân tử DNA kích thước > 4kb, như plasmid pBR322 (4363bp), hoặc SV40 (5243bp). Thí dụ *Sma*I, vị trí xác định chỉ bao gồm C và G (CCC/GGG), hoặc *Xma*III (C/GGCCG) (bảng 2-7). Hiện tượng này sẽ không có trong phân tử nhỏ hơn pBR322 và SV40. Hiện tượng như vậy gọi là tính chuyên biệt của chuỗi mã xác định.

Bảng 2-8 : Sự xuất hiện của vị trí xác định có tính chất tetranucleotide, chứa CpG hoặc GpC dinucleotide ở nhiều phân tử DNA

Enzyme	Sequence	Số vị trí		
		pBR322	ΦX174	SV40
<i>Fnu</i> DII	CGCG	23	14	0
<i>Hha</i> I	GCGC	31	18	2
<i>Hpa</i> II	CCGG	26	5	1
<i>Taq</i> I	TCGA	7	10	1
<i>Hae</i> III	GGCC	22	10	19

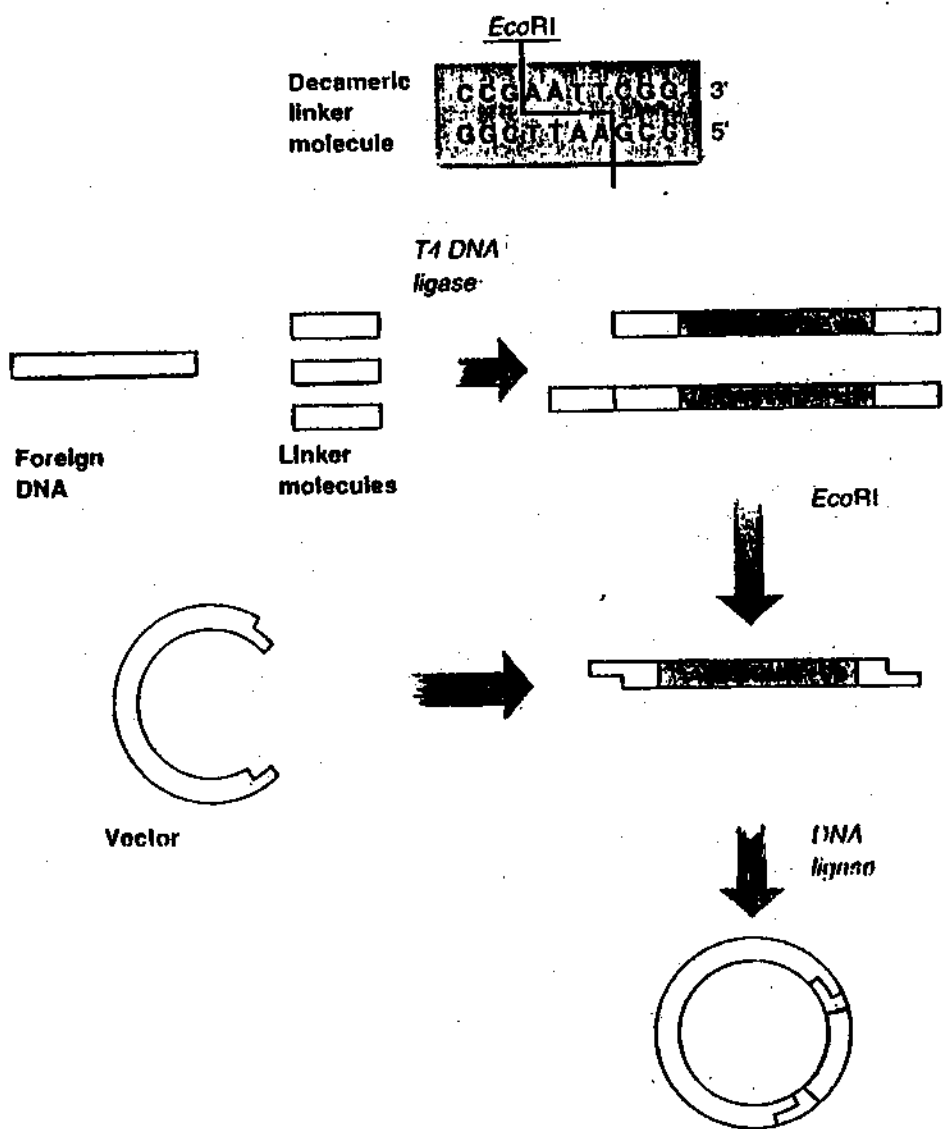
Bảng 2-8 cho thấy enzyme *Fnu*DII (CG/CG) cắt plasmid pBR322 với con số 23 lần, trong khi nó không cắt SV40. Trong khi đó cả hai pBR322 và SV40 bị cắt với cùng tần suất tương tự do *Hae*III (GG/CC). Chuỗi mã xác định này không chứa CpG dinucleotide.

Chuỗi mã dinucleotide CpG cực kỳ hiếm trong DNA của sinh vật eukaryotic.

2.4.6. Cải tiến đầu dây DNA nhờ Enzyme

Hiện có nhiều phương pháp cải tiến đầu dây DNA nhờ restriction enzyme. Việc cải tiến như vậy thường xảy ra, nếu như các đoạn DNA được xử lý có chiều sâu hơn. Nhóm phosphate 5' có thể được loại ra nhờ phản ứng alkaline phosphatase, và có thể được gắn lại tại đầu hydroxylate 5' nhờ một phản ứng xúc tác của T4 polynucleotide kinase (hình 2-9). Phản ứng quấn dây đôi trở lại (annealing) của hai đầu không tiếp hợp, đòi hỏi những đầu có tính chất chồng lấp (overlapping ends), phải được tách ra. Kết quả của hai tiến trình sau cho ra đầu dây theo kiểu blunt end, để rồi nó có thể tiếp tục phản ứng xúc tác do T4 polynucleotide ligase.

Người ta cần phải chuyển đổi đầu dây blunt end thành dây đơn có tính chất protruding (nhô ra ngoài một chút). Điều này có thể được thực hiện một cách hiệu quả nhờ sử dụng λ exonuclease đối với cải tiến đầu 5' (Little và ctv. 1967) và *E. coli* exonuclease III đối với cải tiến đầu 3' (Richardson và ctv. 1964). Phản ứng của dây đôi DNA với exonuclease Bal 31 có thể được dùng cùng một lúc, cho ra các đoạn DNA ngắn hơn, có blunt end ở mỗi đầu dây (Legerski và ctv. 1977).



Hình 2.9 : Phân tử decameric có chứa vùng mục tiêu *EcoRI* gắn với *T4 DNA ligase* ở cả hai đầu của phân tử DNA. Đầu sticky được tạo ra nhờ *EcoRI*. Phân tử DNA này có thể được gắn vào một vectơ, được xử lý cùng một loại restriction endonuclease

Kỹ thuật này cho đến nay vẫn được xem có vai trò quan trọng trong công nghệ recombinant DNA, nhất là trong kỹ thuật DNA sequencing (đọc mã di truyền DNA) được thảo luận ở phần sau. Nó còn được áp dụng trong việc nối lại hai đoạn DNA, hoặc tạo ra những vectơ có tính chất thể hiện (expression vector).

2.4.7. Kết nối các phân tử DNA

Enzyme ngoài chức năng phân cắt, nó còn có chức năng kết nối các phân tử DNA.

2.4.7.1. DNA ligase

E. coli và thực khuẩn thể T4 hình thành một hệ thống mật mã của một enzym, đó là DNA ligase. Phản ứng kết nối (ligation) được hoàn tất tạo ra một thể lai mới giữa vectơ và DNA mục tiêu được gọi là "recombinant". Quần thể của recombinant có thể gia tăng nhờ phản ứng ở nồng độ cao của phân tử DNA (Ferretti và Sgarbetta 1981). Xử lý cho vectơ plasmid ở dạng dây thẳng với alkaline phosphatase, nhằm đẩy ra đầu 5' của gốc phosphate, người ta có thể ngăn ngừa được cả hai hiện tượng : tái lập lại chu kỳ và thành lập dạng dimer của plasmid (hình 2-6 và 2-7).

2.4.7.2. Thể double linker

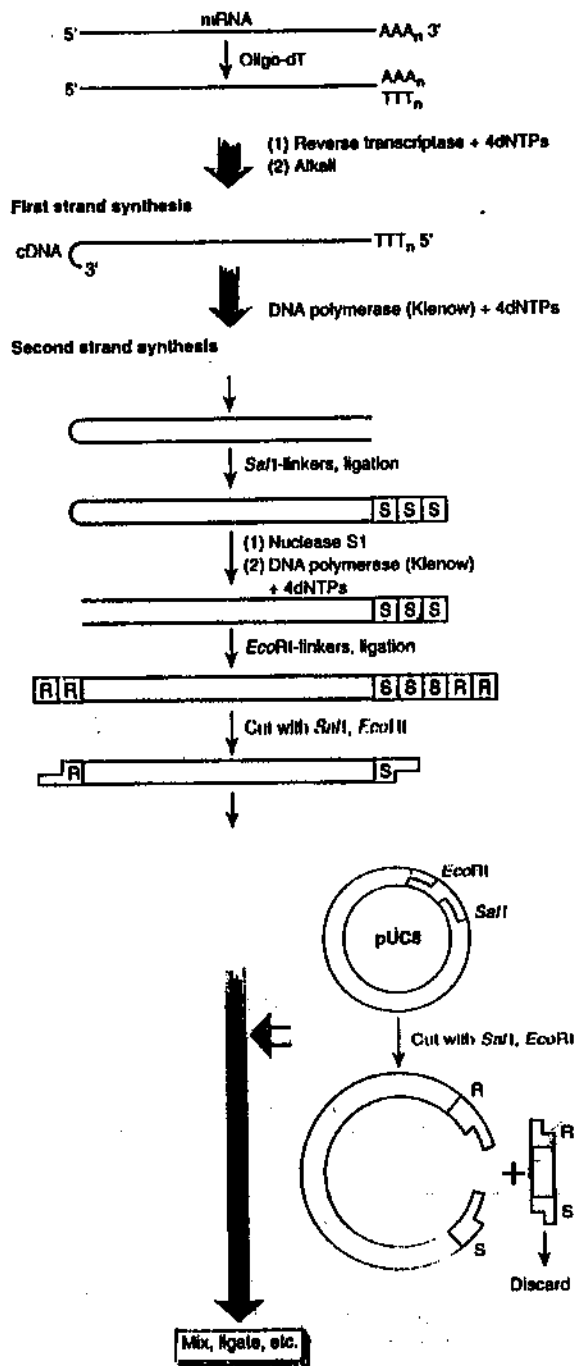
Vectơ plasmid có một loạt vị trí xác định cho kỹ thuật cloning (nhân bản vô tính). Thí dụ như vectơ pUC8. Vectơ này được dùng để tạo dòng vô tính phân tử cDNA dây đôi, nhờ phương pháp có thuật ngữ là "double linker" (Kurtz và Nicodemus 1981, Helfman và ctv. 1983), trong đó, các phân tử linker khác nhau được thêm vào đầu đối diện của cDNA (hình 2-10).

2.4.7.3. Thể adaptor

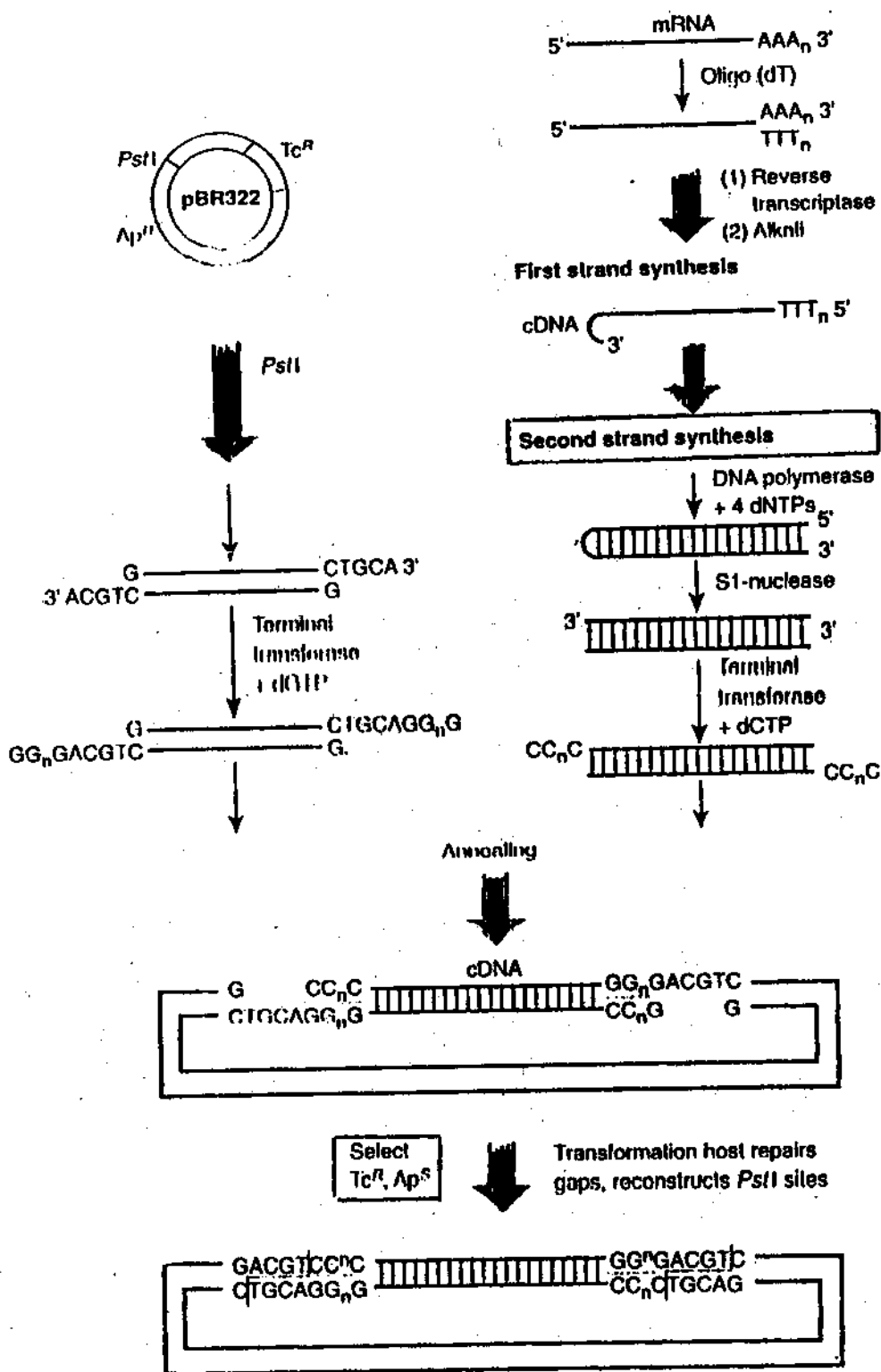
Trong trường hợp restriction enzyme được dùng để tạo ra đầu dây kiểu sticky của thể linker, nó cũng có thể phân cắt DNA mục tiêu tại các vị trí tương ứng bên trong. Như vậy, DNA sẽ được tạo dòng thành hai hoặc nhiều đoạn phụ (subfragment). Người ta đề nghị một giải pháp là chọn một restriction enzyme khác, nhưng không thể nào chọn được khi DNA quá lớn, đòi hỏi nhiều enzyme phục vụ tại nhiều vị trí phân cắt khác nhau. Giải pháp sử dụng methyl hóa vị trí phân cắt hạn chế bên trong với methylase cải tiến cũng được đề nghị. Wu và ctv. (1991) đã tổng hợp được các phân tử adaptor có đầu dây cohesive rất hoàn chỉnh. Xem xét một DNA có đầu dây theo kiểu blunt có vị trí của *Bam*HI bên trong (hình 2-12), nó được tạo dòng với một *Bam*HI-cut vectơ.

2.4.7.4. Tạo đuôi homopolymer

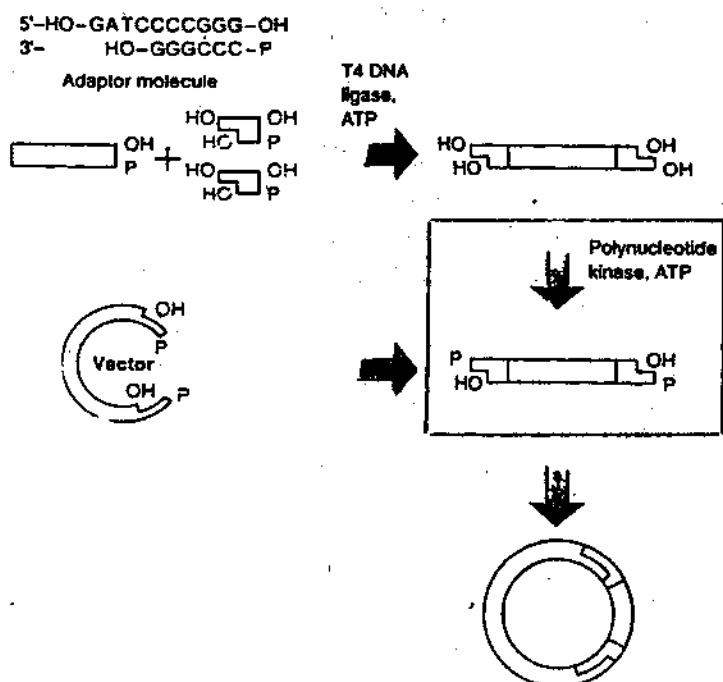
Phương pháp tổng quát để nối các phân tử DNA là sử dụng phản ứng tạo dây đôi ở đầu dây đơn DNA (gọi là annealing) trong những chuỗi mã bổ sung của thể homopolymer. Xong cho thêm vào chuỗi mã oligo dA tại đầu 3' của phân tử DNA, và oligo dT chốt tại đầu 3' của phân tử DNA khác, hai loại hình phân tử này có thể phản ứng annealing để tạo ra thể lai có "chu kỳ dimeric" (hình 2-13). Người ta ứng dụng phương pháp tạo đuôi homopolymer này trong kỹ thuật tạo dòng vô tính cDNA (Smith và ctv. 1979, Old và Primrose 1995). Hình 2-11 là một ví dụ ứng dụng của kỹ thuật này. cDNA dây đôi được chốt đuôi với oligo dC và quần đôi với pBR322, rồi được phân cắt bằng *Pst*I, chốt đuôi bằng oligo dG. Chúng ta sẽ tiếp tục thảo luận trong chương kỹ thuật cloning.



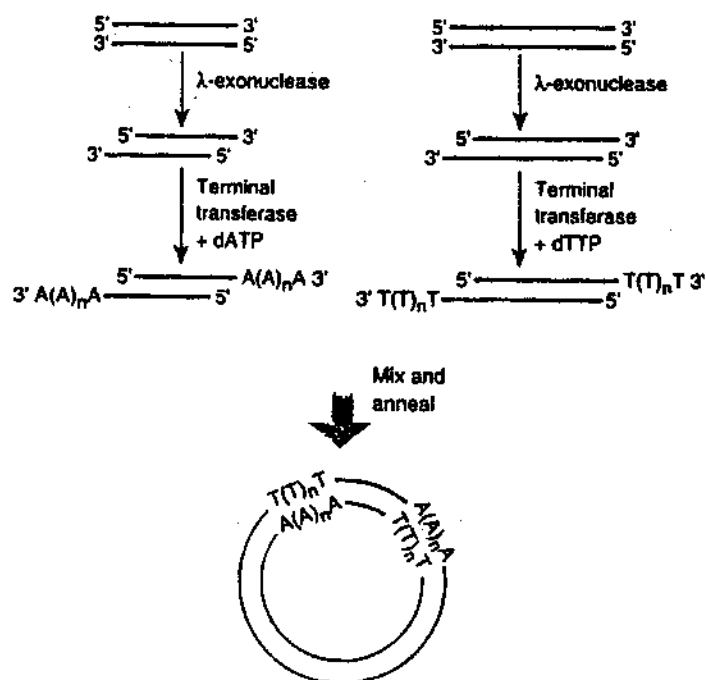
Hình 2.10 : Thể double linker , mRNA được sao ra thành cDNA dây đôi và thể SalI-linker (S) được thêm vào. Cấu trúc vòng hairpin được hình thành bằng cách tự phản ứng priming của dây thứ hai. Sinh tổng hợp dây thứ hai, nhờ nuclease S1, sau đó nó được loại ra nhờ Klenow polymerase. Thể EcoRI linker được gắn thành dây đôi, kiểu đầu dây sticky, được phân cắt nhờ SalI và EcoRI. Phân tử cDNA này cộng với thể linker được phản ứng kết dính nhau thành vectơ được cắt bởi cùng hai enzym.



Hình 2.11 : Tổng hợp một bản sao của cDNA từ mRNA được polyadenyl hóa, và gắn vào một vectơ bằng cách tạo đuôi homopolymer



Hình 2.12 : Sử dụng BamHI adaptor. Một phân tử adaptor được tổng hợp phản ứng kết nối với phân tử DNA. Thể adaptor này được dùng ở dạng 5'hydroxyl để ngăn ngừa hiện tượng tự polymer hóa



Hình 2.13 : Sử dụng enzyme deoxynucleotidyl-transferase của calf-thymus để thêm vào đuôi homopolymer bổ sung của hai phân tử DNA

TÀI LIỆU THAM KHẢO

1. Eun Hyone-Myong. 1996. *Enzymology Primer for Recombinant DNA Technology*. Academic Press (AP). P.1-108.
2. Helfman DM, JR Feramisco, JC Fiddes, GP Thomas, SH Hughes. 1983. *Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library*. Proc. Natl. Acad. Sci. USA 80:31-35.
3. Kurtz DT, CF Nicodemus. 1981. *Cloning of $\alpha_{2\mu}$ globulin cDNA using high efficiency technique for the cloning of trace messenger RNAs*. Gene 13:145-152.
4. Langridge I, P Langridge, PL Bergquist. 1980. *Extraction of nucleic acids from agarose gels*. Anal. Biochem. 103:264-271.
5. Legerski RJ, HB Gray, DL R erson. 1977. *A sensitive endonuclease probe for lesions in deoxyribonucleic acid helix structure produced by carcinogenic or mutagenic agents*. J. Biol. Chem. 252: 8740-8746.
6. Little JW, IR Lehman, AD Kaiser. 1967. *An exonuclease induced by bacteriophage λ . I. Preparation of the crystalline enzyme*. J. Biol. Chem. 242:672-678.
7. Maxam A, W Gilbert. 1980. *Sequencing end-labelled DNA with base-specific chemical cleavages*. Methods in Enzymology 65:499-580.
8. McDonnell MW, MN Simon, FW Studier. 1977. *Analysis of restriction fragments of T7 DNA and determination of molecular weight by electrophoresis in neutral and alkali gels*. J. Mol. Biol. 110:119-146.
9. Mertz JE, RW Davis. 1972 *Cleavage of DNA by RI Restriction endonuclease generates cohesive ends*. Proc. Natl. Acad. Sci. USA 69:3370-3374.
10. Old RM, SB Primrose. 1995. *Principles of Gene Manipulation*. 5th Ed. N.G.CARR. Univ. of Warwick, London, UK.
11. Richardson CC, IR Lehman, A Kornberg. 1964. *A deoxyribonucleic acid phosphatase-exo-nuclease from Escherichia coli*. J. Biol. Chem. 239:251-258.
12. Smith HO, D Nathan. 1973. *A suggested nomenclature for bacterial host modification and restriction systems and their enzymes*. J. Mol. Biol. 81: 419-423.
13. Smith DF, PF Searle, JG William. 1979. *Characterization of bacterial clones containing DNA sequence derived from Xenopus laevis*. Nucleic Acids Res 7:487-506.
14. Tabak HF, RA Flavell. 1978. *A method for the recovery of DNA from agarose gels*. Nucleic Acids Res. 7: 2321-2332.
15. Wu XC, W Lee, L Tran, SL Wong. 1991. *Engineering a Bacillus subtilis expression secretion system with a strain deficient in six extracellular proteases*. J. Bacteriol. 173: 4952-4958.

Chương III

NHỮNG VECTO THÔNG DỤNG TRONG KỸ THUẬT DI TRUYỀN

3.1. PLASMID

Plasmid là những phân tử DNA vòng, xoắn đôi, có tính chất tự lập lại. Nó có ở trong các vi khuẩn, xem như phân tử có tính chất ngoài nhiễm thể (extrachromosomal). Plasmid còn có tên là "supercoiled DNA" do cấu trúc xoắn đôi đặc biệt của nó. Nói chung, tất cả các vi khuẩn đều có plasmid. Một vài plasmid mang thông tin di truyền từ tế bào này sang tế bào khác (gọi là F plasmid); những plasmid khác mang các bộ gen đặc biệt dùng trong những cơ chế không cần thiết, được gọi là "degradative plasmid"; và một vài plasmid không có gen chức năng, có tính tạm thời, được gọi là "cryptic plasmid". Plasmid có qui mô kích thước thay đổi, từ nhỏ hơn 1 đến 500 kb. Mỗi plasmid đều có một chuỗi mã di truyền (sequence) mang chức năng như một nguồn của đặc tính tự tái bản DNA (replication). Nếu không có vị trí nguồn này (*ori*), DNA không thể tái lập lại trong tế bào cây chủ.

Vài plasmid được thể hiện thông qua 10 đến 100 bản sao trong tế bào chủ. Những plasmid này được xem như: plasmid có số bản sao cao. Plasmid khác có 1-4 bản sao / tế bào, được xếp vào nhóm plasmid có số bản sao thấp. Trong trường hợp có hai hoặc nhiều loại plasmid khác nhau không thể ở chung trong cùng một tế bào chủ, người ta gọi đó là plasmid thuộc nhóm đơn, không tương hợp (incompatibility). Những plasmid từ những nhóm không tương hợp như vậy có thể được xử lý cho vào cùng một tế bào. Sự cùng hiện hữu này (coexistence) độc lập với số bản sao của những plasmid riêng biệt. Một vài vi sinh vật được tìm thấy có từ 8 đến 10 plasmid khác nhau. Trong trường hợp này, mỗi plasmid có thể thực hiện nhiều chức năng khác nhau và có bản sao riêng của nó, mỗi thứ thuộc nhóm không tương hợp. Một vài plasmid chỉ có thể tái lập lại trong tế bào của một loài đặc biệt nào đó mà thôi. Trường hợp này được gọi là sự chuyên biệt về nguồn lập lại. Những plasmid khác ít có tính chuyên biệt này, nó có thể tái lập lại trong nhiều loài vi khuẩn. Do đó, người ta xếp loại plasmid thành hai nhóm: nhóm có phổ ký chủ hẹp và nhóm có phổ ký chủ rộng.

Với tính chất tự tái bản, plasmid có một bản chất quan trọng là một vectơ mang DNA đã được tạo dòng (cloned). Tuy nhiên, plasmid thường thiếu nhiều điểm quan trọng khác, cần thiết cho một cloning vectơ có chất lượng tốt. Những đặc điểm ấy là: (1) kích thước nhỏ rất cần thiết, bởi vì hiệu quả của việc chuyển DNA ngoại sinh vào *E. coli* sẽ giảm đi

một cách có ý nghĩa so với plasmid có kích thước lớn hơn 15 kb ; (2) tính đồng nhất (đơn) về các vị trí xác định men endonuclease hoạt động, nơi đó dòng DNA sẽ được gắn vào ; (3) một hoặc nhiều hơn marker di truyền có tính chọn lọc đối với tế bào nhận dòng DNA ngoại sinh. Từ đó, các cloning vector là những plasmid như vậy, sẽ đảm đương công việc được gọi là công nghệ lắp ghép gen.

Một vài tính trạng kiểu hình do gen được plasmid mang :

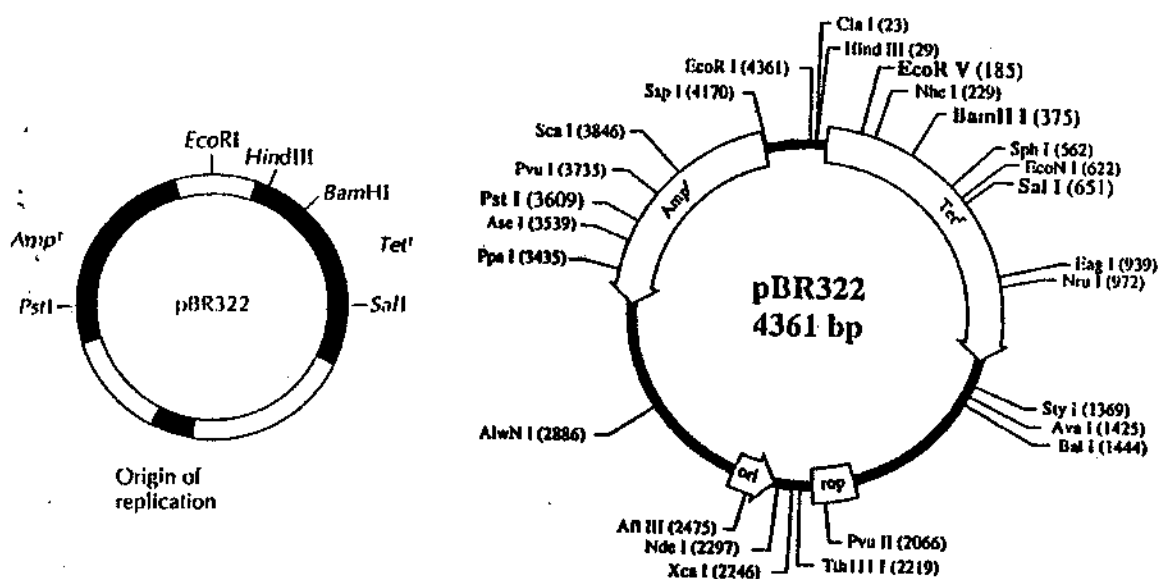
- Tính kháng thuốc kháng sinh
- Sản xuất chất kháng sinh
- Làm giảm các hợp chất thơm
- Sản xuất haemolysin
- Lên men đường
- Sản xuất enterotoxin
- Kháng kim loại nặng
- Sản xuất bacteriocin
- Kích thích các khối u thực vật
- Sản xuất hydrogen sulphide
- Hệ thống R-M được kiểm soát trong sinh vật chủ

Bảng 3-1 : Đặc tính của những conjugative và non-conjugative plasmid của sinh vật có Gram âm

Plasmid	Kích thước (MDa)	Conjugative	Số bản sao của plasmid / nhiễm thể	Kiểu hình
Col E1	4.2	không	10-15	sản xuất Col E1
RFS 1030	5.6	không	20-40	kháng ampicilin
clo DF13	6	không	10	sản xuất cloacin
R6K	25	có	13-38	kháng ampicillin và streptomycin
F	62	có	1-2	
RI	62.5	có	3-6	kháng nhiều loại thuốc
Ent P 307	65	có	1-3	sản xuất enterotoxin

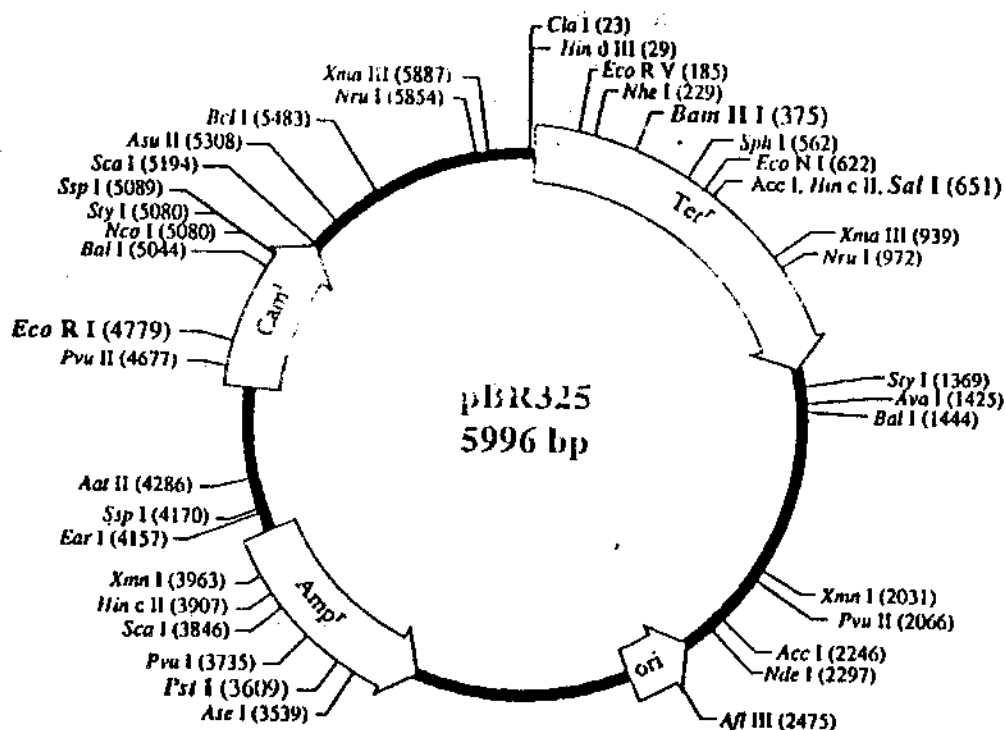
3.1.1. Cloning vector là Plasmid pBR322:

Một trong những nghiên cứu xuất sắc và được ứng dụng rộng rãi là plasmid pBR322. Vectơ tạo dòng này được ký hiệu bằng chữ "p" để chỉ plasmid, ký hiệu BR là tên của F. Bolivar và R. Rodriguez - hai nhà nghiên cứu tìm ra plasmid này, và số 322 là ký hiệu của nhà nghiên cứu đặt ra. Plasmid pBR322 có qui mô về kích thước là 4361 bp. Nó mang hai gen kháng với thuốc kháng sinh: một đối với ampicilline (ký hiệu là *Amp^r*), một đối với tetracycline (ký hiệu là *Tet^r*). Plasmid này cũng có vị trí đồng nhất của *Bam*HI, *Hind*III, và *Sal*I ở trong gen *Tet^r*, vị trí *Pst*I ở trong gen *Amp^r*. Vị trí *Eco*RI nằm ở chỗ không có tín hiệu di truyền nào. Nguồn của sự tự tái bản (origin of replication) chỉ có chức năng trên *E. coli*. Số bản sao rất cao của *E. coli* được duy trì, và không ở trạng thái sẵn sàng để chuyển sang bacteria khác (hình 3-1 và 3-2).



Hình 3.1 : Bản đồ di truyền của plasmid pBR322, bao gồm những marker kháng thuốc kháng sinh *Amp^r* (kháng ampicillin), *Tet^r* (kháng tetracycline), nguồn tự tái bản (*ori*), một gen kiểm soát số bản sao của plasmid (*rop*), những vị trí xác định của *Hind*III, *Sal*I, *Bam*HI, và *Pst*I.

Làm thế nào để pBR322 hoạt động như một cloning vectơ? Phân tử pBR322 là một DNA vòng kín, thuần khiết. pBR322 được cắt với một restriction enzyme, tại nơi có chứa gen kháng với thuốc kháng sinh, tạo ra DNA dây đơn, thẳng, dạng sticky end (đầu dây so le khác với dạng blunt end là đầu dây bằng nhau). Phân tử dây thẳng này được kết hợp với DNA của một nguồn lạ nào đó. DNA này được cắt ra bằng một restriction enzyme giống như enzyme đã được sử dụng trên plasmid, để có hai đầu thuộc dạng sticky end tương ứng (như ngàm của hai thanh gỗ khớp với nhau). Hỗn hợp DNA này được xử lý bằng men T4 DNA ligase, với sự có mặt của ATP. Trong điều kiện như vậy, hàng loạt các phối hợp khác nhau sẽ được tạo ra, chứa đựng cả plasmid DNA vòng nguyên thủy. Để giảm thiểu các liên



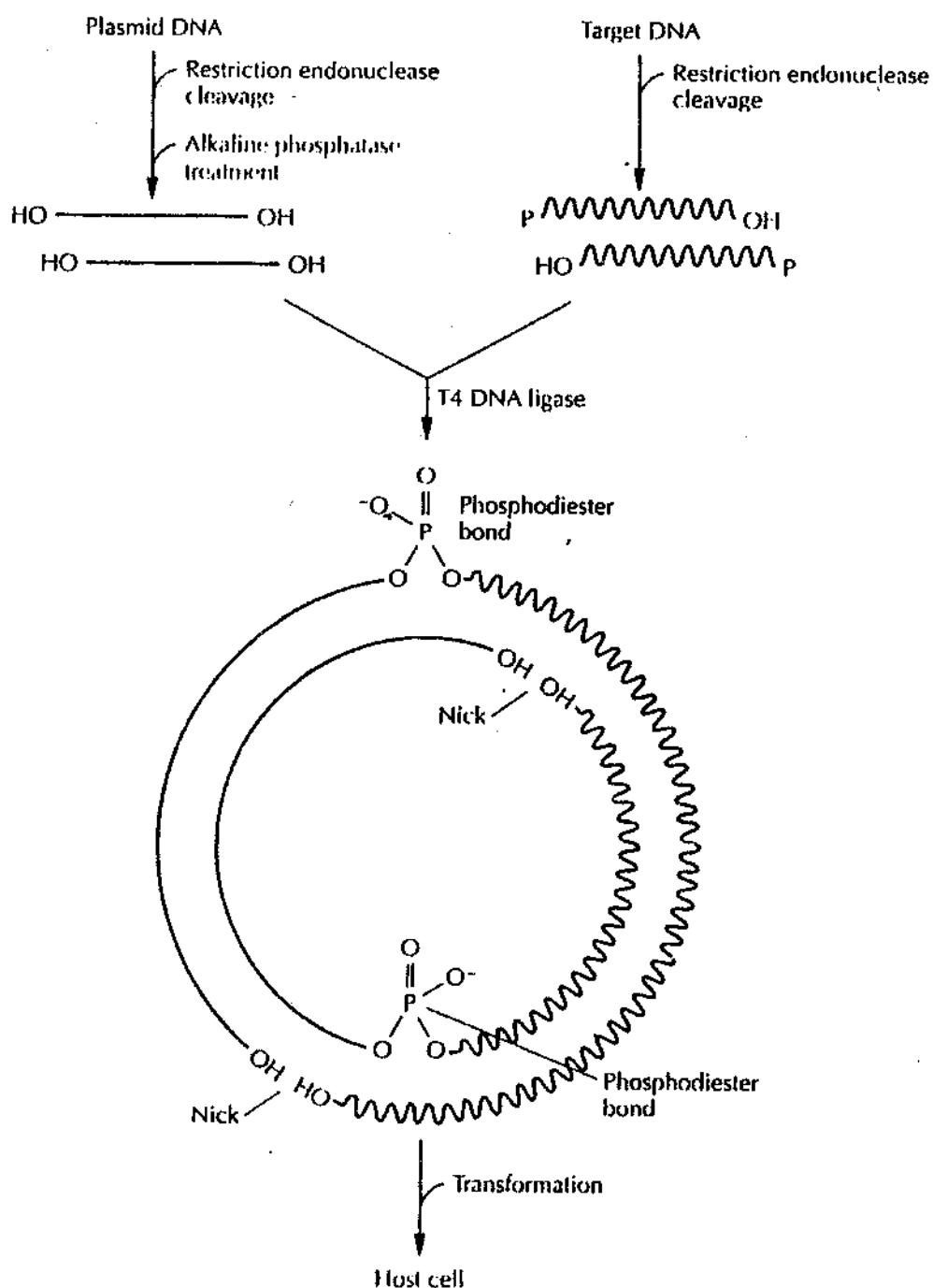
Hình 3.2 : Bản đồ di truyền của plasmid pBR325 có độ dài 55996 bp lớn hơn pBR322, chứa các gen *mp^r* (kháng ampicillin), *Cam^r* (kháng chloramphenicol), và *Tet^r* (kháng tetracycline)

kết không mong muốn, plasmid có gắn DNA mới sẽ được xử lý bằng enzyme alkali phosphatase, nhằm loại ra nhóm phosphate 5' từ DNA plasmid dây thẳng. Nhờ vậy, T4 DNA ligase không thể gắn với đầu của dây thẳng DNA plasmid đã được phản phosphoryl hóa (dephosphorylated). Tuy nhiên, hai cầu nối phosphodiester được tạo ra bởi T4 DNA ligase sau phản ứng nối kết và tạo vòng (circularization) của DNA đã xử lý alkaline phosphatase, đủ sức giữ cả hai phân tử dính nhau lại (DNA mới và plasmid), cho dù có sự hiện diện của hai nick (nối gốc OH và OH). Sau khi chuyển nạp xong, các nick sẽ được hàn lại bởi hệ thống DNA ligase của tế bào chủ (hình 3-3).

3.1.2. Đặc tính cơ bản của plasmid

Plasmid phân bố rất rộng trong sinh vật bậc thấp (prokaryotes), kích thước thay đổi từ 1×10^6 daltons đến 200×10^6 daltons.

Một vài tính trạng về kiểu hình do gen mang plasmid: kháng thuốc kháng sinh, sản xuất ra chất kháng sinh, phân hóa hợp chất có mùi thơm, sản xuất hemolysin, lên men đường, sản xuất enterotoxin, kháng với kim loại nặng, sản xuất bacteriocin, kích thích khối u thực vật, sản xuất hydrogen sulphide.



Hình 3. 3 : Cloning phân tử DNA mục tiêu vào một vectơ plasmid. Sau khi phân cắt bằng restriction endonuclease và xử lý alkaline phosphatase, plasmid DNA được kết nối với các enzyme phân cắt hạn chế, hai đến bốn nick được hàn kín lại (nick là chỗ hở của hai gốc OH phản ứng với nhau). Hai phân tử gắn với nhau theo kiểu đồng hóa trị. Sau khi du nhập vào tế bào chủ, chu kỳ tự tái bản phát triển tạo ra các phân tử DNA vòng, không có nick.

Cryptic plasmid là plasmid có tính trạng kiểu hình chưa được biết.

Plasmid có thể được chia ra làm hai loại : conjugative(tiếp hợp) và non-conjugative (không tiếp hợp) tùy thuộc vào nó có mang hay không mang gen chuyển nạp, gọi là *tra* gen, gen này kích hoạt sự phân cắt tế bào của vi khuẩn.

Plasmid cũng được chia trên cơ sở số bản sao (copy) trên tế bào nhiều : relaxed plasmid, và ít, hoặc bị giới hạn : stringent plasmid.

Conjugative plasmid thường có trọng lượng phân tử cao, xuất hiện một đến ba copy/nhiễm thể. Trong khi non-conjugative plasmid có trọng lượng phân tử thấp, xuất hiện nhiều copy/nhiễm thể.

3.1.3. Sự thuần khiết của plasmid DNA

Điều kiện tiên quyết cho kỹ thuật nhân bản vô tính (cloning) là sự thuần khiết của plasmid DNA. Đó là vấn đề thuộc về phương pháp. Điều kiện lý tưởng có được khi mỗi tế bào bị vỡ, đủ để cho phép plasmid DNA thoát ra, mà không bị tạp quá nhiều DNA ở nhiễm thể. Các DNA nhiễm thể có phân tử lớn sẽ được loại bỏ, nhờ ly tâm với vận tốc lớn.

Phương pháp Vinograd : ly tâm trong dung dịch CsCl, có chứa ethidium bromide (EtBr).

Phương pháp Birnboim và Doly : quan sát trong phổ pH rất hẹp (12.0 - 12.5) trên dây đơn DNA thẳng.

3.1.4. Đặc tính cần thiết của plasmid trong kỹ thuật cloning

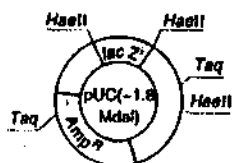
Có ba đặc tính :

- Trọng lượng phân tử thấp
- Khả năng thăm dò các tính trạng chọn lọc trên tế bào chủ
- Có cloning site đủ cho một lượng lớn restriction enzyme

3.1.5. Những plasmid khác trong kỹ thuật cloning

Plasmid pBR322 được xem dùng vectơ trong kỹ thuật cloning rất phổ biến. Nó chỉ có một ít vị trí cloning có tính đồng nhất (cloning site), việc sử dụng nó cần một qui trình có tính chọn lọc về thời gian. Vì vậy, người ta cần phải phát triển thêm những hệ thống khác. Thí dụ plasmid pUC19 có 2686 bp, chứa gen kháng ampicilline, β -galactosidase gene (*lacZ'*) trong cơ chế operon về lactose của *E.coli*, gen *lacI* sản xuất ra repressor protein - điều tiết sự thể hiện của *lacZ'*, một đoạn mã ngắn có nhiều cloning site (như *EcoRI*, *SacI*, *KpnI*, *XmaI*, *BamHI*, *XbaI*, *Sall*, *HincII*, *AccI*, *BspMI*, *PstI*, *SphI*, *HindIII*), và nguồn của sự tái lập lại từ pBR322 (hình 3-4).

pUC19	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8	
	THR	MET	ILE	THR	pro	ser	leu	his	ala	cys	arg	ser	thr	leu	glu	asp	pro	arg	val	pro	ser	ser	ASN	SER	LEU	ALA	
	ATG	ACC	ATG	ATT	ACG	CGA	AGC	TTG	CAT	GCC	TGC	AGG	TGG	ACT	CTA	GAG	GAT	CCC	CGG	GTA	CCG	AGC	TCG	AAT	TCA	CTG	GCC
					HindIII		SphI		PstI		SalI	AccI	HincII		XbaI		BamHI		XmaI	SmaI		KpnI	SacI	EcoRI		HaeIII	
pUC18	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	7	8	
	THR	MET	ILE	THR	ASN	SER	ser	ser	val	pro	gly	asp	pro	leu	glu	ser	thr	cys	arg	his	ala	ser	leu	ala	LEU	ALA	
	ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	GTA	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGG	CAT	GCA	AGC	TTG	GCA	CTG	GCC
					EcoRI		SacI		KpnI		XmaI		BamHI		XbaI		SalI		PstI		SphI		HindIII			HaeIII	
pUC13	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	5	6	7	8			
	THR	MET	ILE	THR	pro	ser	leu	gly	cys	arg	ser	thr	leu	glu	asp	pro	arg	ala	ser	ser	ASN	SER	LEU	ALA			
	ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	GGC	TGC	AGG	TGG	ACT	CTA	GAG	GAT	CCC	CGG	GCG	AGC	TCG	AAT	TCA	CTG	GCC		
					HindIII				PstI		SalI	AccI	HincII		XbaI		BamHI		XmaI	SmaI		SacI	EcoRI		HaeIII		
pUC12	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18			
	THR	MET	ILE	THR	ASN	SER	ser	ser	pro	gly	asp	pro	leu	glu	ser	thr	cys	ser	pro	ser	leu	ala	LEU	ALA			
	ATG	ACC	ATG	ATT	ACG	AAT	TCG	AGC	TCG	CCC	GGG	GAT	CCT	CTA	GAG	TCG	ACC	TGC	AGC	CCA	AGC	TTG	GCA	CTG	GCC		
					EcoRI		SacI			XmaI		BamHI		XbaI		SalI		PstI		HindIII				HaeIII			
pUC9	1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	5	6	7	8								
	THR	MET	ILE	THR	pro	ser	leu	ala	ala	gly	arg	arg	thr	pro	gly	ASN	SER	LEU	ALA								
	ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	GCT	GCA	GGT	CGA	CGG	ATC	CCC	GGG	AAT	TCA	CTG	GCC							
					HindIII				PstI		SalI		BamHI		XmaI		EcoRI		HaeIII								
pUC8	1	2	3	4	5	6	1	2	3	4	5	6	7	8	9	10	11	7	8								
	THR	MET	ILE	THR	ASN	SER	arg	gly	ser	val	asp	leu	gln	pro	ser	leu	ala	LEU	ALA								
	ATG	ACC	ATG	ATT	ACG	AAT	TCC	CGG	GGA	TCC	GTC	GAC	CTG	CAG	CCA	AGC	TTG	GCA	CTG	GCC							
					EcoRI		XmaI		BamHI		SalI	AccI	HincII		PstI		HindIII		HaeIII								
pUC7	1	2	3	4	5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	6							
	THR	MET	ILE	THR	ASN	ser	pro	asp	pro	ser	thr	cys	arg	ser	thr	asp	pro	gly	asn	SER							
	ATG	ACC	ATG	ATT	ACG	AAT	TCC	CCG	GAT	CCG	TCG	ACC	TGC	AGG	TCG	ACG	GAT	CCG	GGG	AAT	TCA						
					EcoRI			BamHI		SalI	AccI	HincII		PstI		SalI	AccI	HincII		EcoRI							



Hình 3.4 : Bản đồ di truyền của những plasmid pUC. Vị trí cloning theo kiểu đa dòng (MCS) được gắn vào gen *lacZ'*, nhưng không can thiệp vào chức năng của gen. Những codon bổ sung xuất hiện trong gen *lacZ'* là kết quả của thể polylinker. Những vùng của polylinker này (MCS) rất đồng nhất với các vectơ M13mp.

Quy trình chọn lọc của pUC19: tế bào mang pUC19 chưa được cải tiến được phát triển trong môi trường isopropylthiogalactoside (IPTG) - một inducer trong *lac* operon. Sản phẩm của gen *lacI* không thể kết dính với vùng hoạt động của promoter - operator của gen *lacZ'*, cho nên gen *lacZ'* trong plasmid được chuyển mã và giải mã. Protein *lacZ'* phối hợp với protein đã được mang tín hiệu di truyền của DNA nhiễm thể, để tạo ra một β -galactosidase lai và hoạt động mạnh mẽ. Cuối cùng chất còn lại 5-bromo-4-chloroindolyl- β -galactosidase (X-gal) trong môi trường, sẽ bị thủy phân bởi β -galactosidase lai này, cho ra một sản phẩm màu xanh lơ. Trong điều kiện như vậy, các colonies có chứa pUC19 chưa cải tiến sẽ xuất hiện màu xanh này.

Trong thí nghiệm tạo dòng bằng pUC19, phân tử DNA của sinh vật (source DNA) được cắt bằng restriction endonuclease để có điểm tương ứng cho kỹ thuật cloning sequence. Phân tử DNA này được kết với pUC19. Người ta xử lý pUC19 bằng restriction endonuclease giống như đối với DNA. Sau đó, người ta cho xử lý với alkaline phosphatase. Sau khi gắn chúng vào bằng T4 DNA ligase, phản ứng kết hợp được thực hiện, do sự chuyển nạp vào tế bào chủ. Tế bào chủ có thể tự tổng hợp một phần β -galactosidase để kết hợp với sản phẩm của *lacZ'* gen, tạo ra enzym chức năng. Tế bào chủ này được đặt sang môi trường có chứa ampicillin, IPTG, và X-gal. Những tế bào không được chuyển nạp, không thể phát triển trong môi trường có ampicillin. Tế bào có plasmid được tái lập chu kỳ, sẽ phát triển được trong môi trường ampicillin, bởi vì nó có thể tạo ra β -galactosidase chức năng, nó sản xuất ra các colonies có màu xanh lơ. Trái lại, tế bào chủ mang cấu trúc plasmid-cloned DNA sẽ cho ra colonies màu trắng, trên môi trường này. Bởi vì DNA được gắn vào vị trí restriction endonuclease trong nhiều chuỗi mã của dòng sẽ đột phá khả năng đọc của gen *lacZ'* và ngăn cản việc sản xuất protein *lacZ'* chức năng, tạo ra hệ quả là không có một β -galactosidase lai nào được sản xuất. Sự vắng mặt β -galactosidase làm cho X-gal trong môi trường không chuyển đổi thành hợp chất màu xanh. Colonies màu trắng (positive) sẽ được thanh lọc ngay sau đó, để xác định xem cái nào mang chuỗi mã DNA cần thiết cho nghiên cứu.

Bổ sung với pBR322 và các series của pUC plasmid, nhiều cloning vectơ khác cũng đã được xem xét. Nguyên tắc chung bao gồm hai yêu cầu cơ bản như sau cho kỹ thuật recombinant DNA là:

- Chọn cloning site
- Tìm cách nào dễ nhất để xác định tế bào có cấu trúc plasmid-cloned DNA

Vị trí của restriction endonuclease là chức năng cực kỳ quan trọng trong nghiên cứu recombinant DNA.

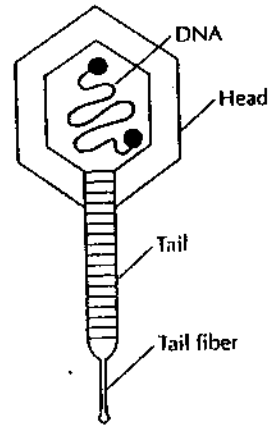
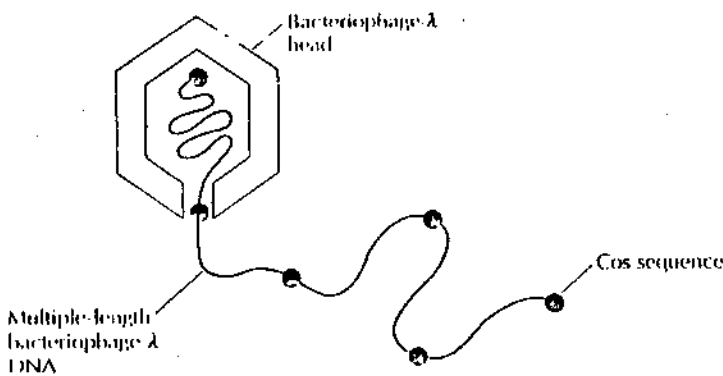
3.2. VECTƠ LAMBDA CỦA THỰC KHUẨN THỂ

Những vectơ là plasmid được sử dụng trong kỹ thuật cloning DNA thường mang trên mình nó DNA có kích thước 10kb. Tuy nhiên để tạo ra thư mục, người ta cần duy trì những dòng DNA có kích thước lớn hơn. Nhằm thỏa mãn yêu cầu này, người ta đã sử dụng thể phage của virus trên *E. coli* (bacteriophage) gọi là lambda (λ), trong kỹ thuật cloning (hình 3-5).

3.2.1. Nguyên tắc hoạt động

Trong chu kỳ sống của nó, bacteriophage λ xâm nhiễm vào *E. coli*, đưa vào đó DNA của siêu vi, do đó sẽ có hai khả năng xảy ra. Thực khuẩn thể λ có thể xâm nhập vào chu kỳ phân hủy tế bào, sau 20 phút, nó làm cho tế bào chủ bị phân hủy, rồi phóng thích ra hàng trăm phân tử phage. Luân phiên như vậy, DNA của thực khuẩn thể λ có thể nhập vào

A



Hình 3.5 : Phân tử DNA của thực khuẩn thể λ

- (A) Tự tái bản DNA từ dạng vòng của λ làm cho nó trở thành dạng dây thẳng, hình thành các đoạn chồng lấp lên nhau, với độ dài khoảng 50 kb.
- (B) Mỗi đầu đều chứa đầy 50 kb đơn vị của λ DNA, trước khi đuôi được tổng hợp gắn vào sau

nhằm thể của *E. coli* như một prophage và duy trì tình trạng này như một yếu tố phân giải (lysogeny). Tuy nhiên, trong điều kiện bị stress do dinh dưỡng hoặc môi trường, DNA của thực khuẩn thể λ này có thể bị chia cắt ra, rồi đi vào chu kỳ phân giải. DNA của thực khuẩn thể λ có qui mô khoảng 50 kb về chiều dài, trong đó, ước khoảng 20kb cần và đủ cho quá trình hợp nhất - phân cắt (I/E viết tắt từ integration/excision). Để tạo ra thư mục của genome, người ta yêu cầu 20kb của DNA có thể được thay bằng 20 kb của cloned DNA. Phân tử DNA này có thể được duy trì như một dạng tái tổ hợp (recombinant) của DNA thực khuẩn thể λ trong suốt chu kỳ phân giải có tính chất bất buộc.

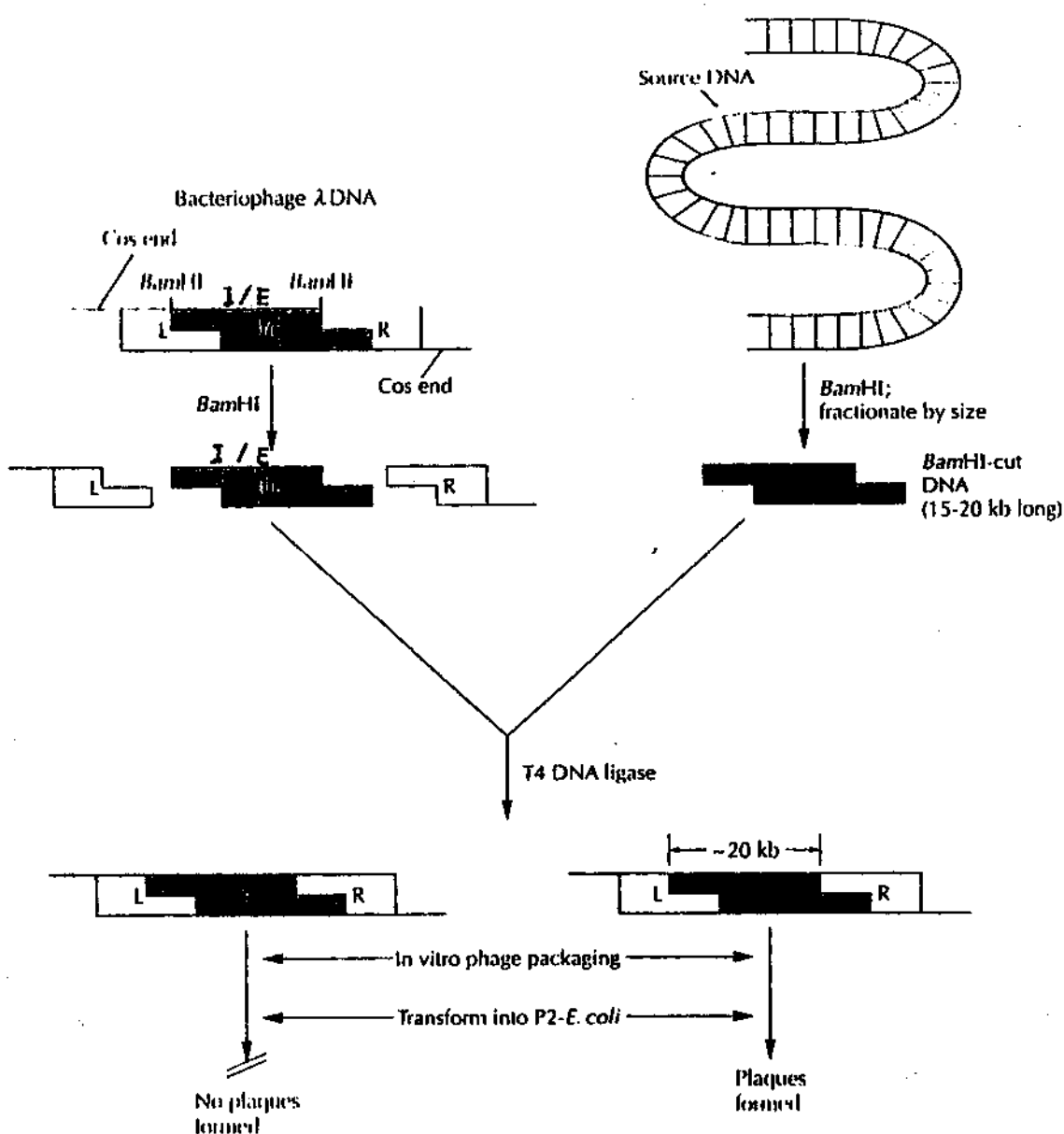
Làm thế nào thực khuẩn thể λ có chức năng trong hệ thống cloning, người ta cần phải xem xét nó dưới góc độ sinh học phân tử của chu kỳ phân giải. Thực khuẩn thể λ gồm có một đuôi protein dạng hình ống, dài, với một sợi protein ở đuôi và một đầu protein có qui mô DNA là 50kb. Sản xuất và tổng hợp phần đầu và đuôi, cũng như quá trình bao kín DNA là kết quả của chuỗi mã có tính dung hợp cao (highly coordinated sequence). DNA ở phần đầu của λ có qui mô 50kb, dây thẳng, với một khúc phát triển thêm, dây đơn, 12 base, dính ở vị trí 5'. Đoạn phát triển thêm này được gọi là đầu dính [cohesive (cos) ends]. Bởi vì nó chứa chuỗi mã có tính bổ sung cho mỗi chuỗi khác. Sau khi chèn λ DNA vào *E. coli* thông qua đuôi, cặp base của cos end này sẽ nối lại hình thành phân tử DNA vòng. Trong pha đầu của chu kỳ phân giải, sự tự tái bản DNA từ phân tử vòng sẽ tạo ra một dạng thẳng

λ DNA, nó bao gồm nhiều độ dài tiếp cận nhau, với 50kb đơn vị. Mỗi đầu mới tổng hợp này được trám đầy 50kb đơn vị của DNA. Sau đó, đuôi được gắn vào, hoàn thành quá trình tạo nên một phân tử xâm nhiễm mới của virus. Thể tích của đầu thực khuẩn thể λ đủ sức chứa 50kb. Nếu DNA có ít hơn 38kb được gắn vào đầu thực khuẩn thể λ , thì phage được tạo ra sẽ có tính chất không lây nhiễm (non-infective), ngược lại, nếu nó quá lớn (> 52 kb), thì đầu λ không chứa nổi. Vị trí của cos sequence dính rả rác trên dây thẳng 50kb của λ DNA, đảm bảo mỗi đầu chỉ nhận chính xác một lượng DNA nào đó. Định vị gắn của của đầu là một enzym ghi nhận cos sequence của dây đôi, và cắt DNA ở tại vị trí này, khi DNA được gắn vào đầu.

Bên cạnh việc xác định thực khuẩn thể λ dung hợp như thế nào, người ta còn nghiên cứu in vitro hệ thống đóng kín DNA vào λ . Trộn hai đầu hoàn toàn trống rỗng với nhau, thực khuẩn thể λ (50kb), đuôi, phân tử gây nhiễm có thể được tạo ra trong ống nghiệm.

Một trong những vectơ thuộc thực khuẩn thể λ có hai *Bam*HI sites để tác động vùng I / E. Khi DNA thuần khiết của thực khuẩn thể được cắt với *Bam*HI, sẽ có ba đoạn được tạo thành. Vùng bên trái (L) chứa thông tin di truyền để sản xuất ra đầu và đuôi. Vùng bên phải (R) mang gen giúp DNA tự tái bản và phân hủy tế bào. Vùng giữa (I / E) có gen cho quá trình tổng hợp và phân cắt. Mục tiêu của công nghệ di truyền là thay vùng giữa của λ DNA bằng cloned DNA, ước khoảng 20 kb về chiều dài. DNA của genome được cắt bằng *Bam*HI. Người ta phân lập các đoạn DNA từ 15 đến 20 kb. Hai mẫu DNA này được kết hợp lại và định ôn với men T4 DNA ligase. Đầu của thực khuẩn thể còn trống và đuôi được thêm vào. Trong điều kiện như vậy, DNA với 50kb đơn vị sẽ được đóng kín vào đầu λ , như vậy thực khuẩn thể có tính lây nhiễm đã được hình thành. Những sản phẩm khác của phản ứng dung hợp này có thể không được đóng lại vì nó quá lớn (> 52 kb) hoặc quá nhỏ (< 38 kb). Thực khuẩn thể λ có tính chất tái tổ hợp chỉ chịu trải qua chu kỳ phân giải trong *E. coli*, nhưng không cho phép thực khuẩn thể λ tái lập lại vùng tiếp xúc I / E để phát triển. Thực khuẩn thể λ có tính tái tổ hợp được duy trì bằng cách phát triển liên tục trên *E. coli* sống (hình 3-6).

Kho lưu trữ của thực khuẩn thể λ có thể được thanh lọc bằng cả hai phương pháp: DNA probe và xét nghiệm miễn dịch học. Trong hệ thống vectơ plasmid, colonies của vi khuẩn được dùng để trắc nghiệm, nhưng trong hệ thống vectơ thực khuẩn thể, các khu vực phân giải (lysis) có thực khuẩn thể được xem xét như matrix và được thực hiện chức năng. Đối với kỹ thuật lai DNA, protein của thực khuẩn thể được lấy ra và DNA lai được tách thành dây đơn, quấn với nhau dạng matrix. Đối với xét nghiệm miễn dịch học, các protein được mã hóa bởi gen của dòng vô tính, sẽ được tổng hợp trong quá trình phân giải. Những protein này được chuyển đến vùng phân giải (plaque), rồi tạo thành matrix.



Hình 3.6 : Hệ thống cloning của thực khuẩn thể λ . Lambda được xử lý để tạo ra hai vị trí của BamHI, nằm kề vùng I/E (tổng hợp/phân cắt) của genome thực khuẩn thể. Sự kéo dài làm hình thành đuôi cos cụ λ DNA để cloning. DNA nguồn sẽ được cắt bởi BamHI và được phân đoạn theo kích thước (15 - 20 kb). Thực khuẩn thể λ cũng tự phân cắt bằng BamHI. Hai mẫu DNA này được trộn vào nhau và cho xử lý bằng T4 DNA ligase. Phản ứng kết nối xảy ra, sẽ chứa một khối lượng DNA khác, bao gồm : thực khuẩn thể λ mới, vùng L và R của λ với 20 kb của DNA thay vào vùng I/E. Những phân tử này nằm gói gọn trong đầu của λ in vitro. Và những phân tử infective được hình thành sau khi nó được thêm đuôi.

3.2.2. Những vectơ lambda thực khuẩn thể cải tiến

Những vectơ lambda thực khuẩn thể cải tiến nhằm mục đích :

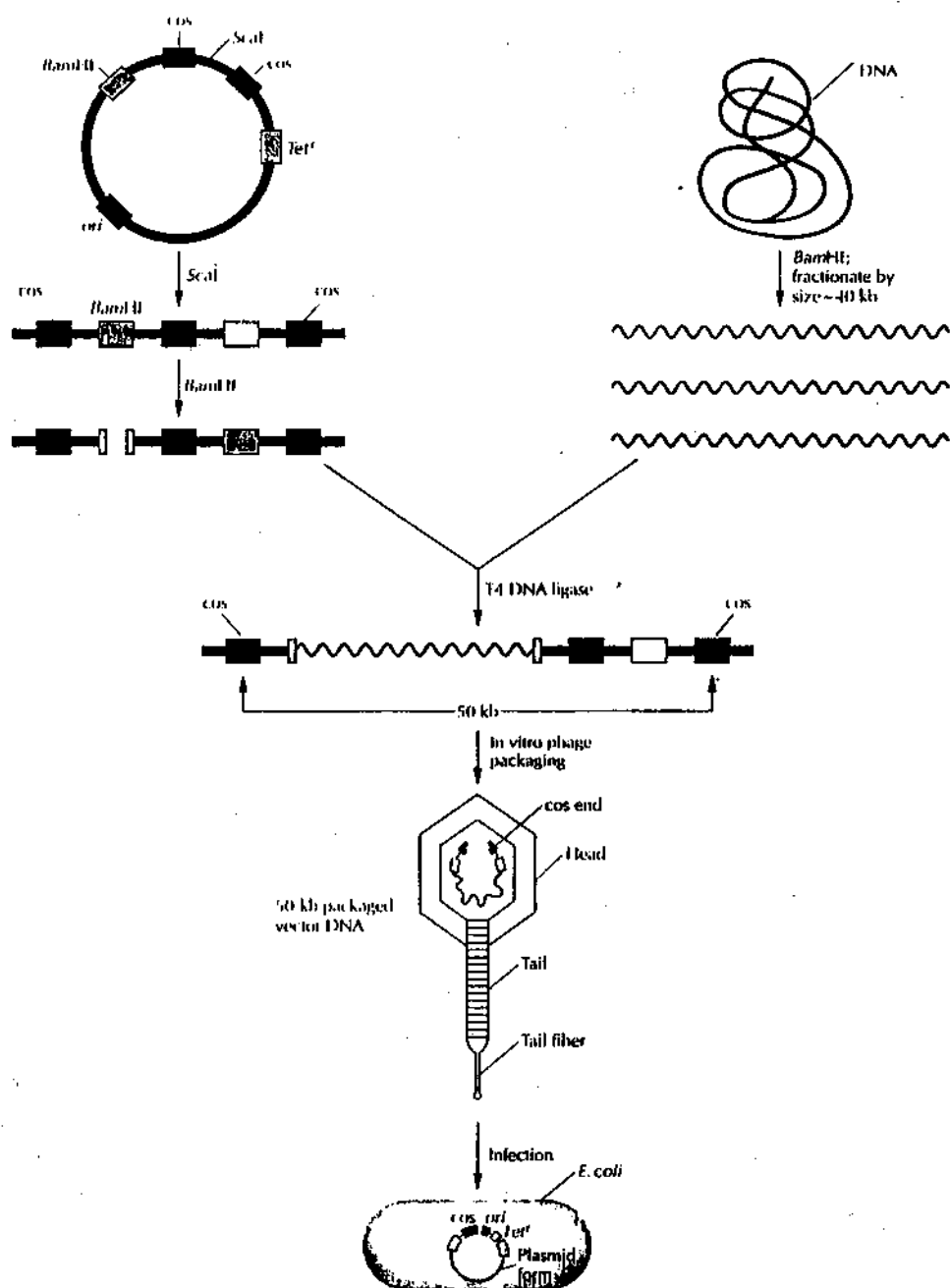
- Gia tăng khả năng của các đoạn DNA thích hợp với nhiều restriction enzyme.
- Chia ra nhiều phương pháp để có thể chọn lựa được cách thức recombinant mà mình cần có.
- Cho phép có được các RNA probe sẵn sàng cho việc chuyển mã của DNA lạ gắn vào vectơ. Điều này giúp cho chúng ta dễ dàng thanh lọc các thư mục trong tiến trình chromosome walking (sẽ được thảo luận ở các phần sau), thí dụ như λ ZAP.
- Phát triển những vectơ trong tiến trình gắn vào cDNA của sinh vật bậc cao, dưới hình thức một polypeptide dung hợp, với β -galactosidase. Hình thức này rất hữu dụng trong việc chọn lọc kháng thể, thí dụ như vectơ λ gt11.

Thế hệ gần đây nhất của vectơ λ kết hợp với khả năng to lớn của DNA lạ mà nó mang, tiến gần đến hạn chế về lý thuyết là 23kb. Những vectơ được thay thế là EMBL3 và EMBL4 (Frischaud và ctv. 1983) đã cung cấp phương tiện tốt cho chuỗi mã có tính chất polylinker, nằm gần các đoạn có thể thay thế được. (Chú thích: linker là chuỗi mã DNA nhân tạo, bổ sung vào các đoạn phân tử DNA mục tiêu, nhằm tạo ra một vị trí phân cắt hạn chế của enzyme restriction site, còn polylinker được hiểu như là phân tử có nhiều linker hợp lại). Các phage với những thể insert đính bên trong nó; có thể được chọn lọc bằng kiểu hình Sp^+ , chính là chi^+ . Thể cải tiến từ EMBL3 có những đột biến EMBL3 *Sam*, EMBL3 *Aam Sam*. Những đột biến trong vectơ λ không chỉ làm gia tăng khả năng chứa của nó, mà còn được sử dụng trong hệ thống chọn lọc những DNA sequence phân lập, liên kết với các gen suppressor.

3.3. COSMID

Vectơ cosmid trong kỹ thuật cloning được khởi xướng từ 1978 do Collin và Hohn. Nó đã trở thành vectơ chính cho cloning các DNA có kích thước lớn hơn 20 kb, và nhỏ hơn 50 kb, mãi cho đến khi YAC ra đời vào năm 1987 (Burke và ctv. 1987). Cosmid về căn bản chính là plasmid vectơ, nhưng nó có chứa một hoặc hai lambda cos site, cho phép DNA gắn vào trong phân tử lambda virus. Đặc điểm này cho phép sự chuyển có hiệu quả DNA ở dạng capsulated vào *E. coli* do phản ứng transfection.

Cosmid được sử dụng rất sớm trong công nghệ sinh học thực vật, và hiện còn giá trị như một cloning vectơ, có qui mô 40 kb. Về cơ bản, nó là plasmid vectơ, nó có chứa thêm lambda cos site, cho phép nó có thể mang DNA vào bằng kỹ thuật in vitro. Người ta cải tiến cosmid vectơ thành vectơ chuyển nạp gen ở thực vật



Hình 3.7 : Hệ thống cloning của cosmid. Nó có chứa *ori* của *E. coli* (cho phép cosmid được duy trì như plasmid trong *E. coli*). Hai *cos* site gần với vị trí phân cắt của *Scal*, *Bam*HI. DNA nguồn bị cắt bởi *Bam*HI để phân đoạn DNA có kích thước 40kb. Gen kháng tetracycline (*Tet^r*) định vị gần *cos* site. Phân tử plasmid DNA được cắt bởi *Bam*HI và *Scal*. Hai mẫu DNA này được trộn vào nhau và được kết nối bằng T4 DNA ligase. Sau khi kết nối xong, những phân tử này được gói gọn trong phần đầu của λ , và những phân tử infective sẽ được hình thành sau khi tạo đuôi.

Vectơ được tiêu hóa, phân ra làm hai nhánh hoặc hai cánh tay (arm). Trước hết, vectơ được tiêu hóa bằng *Hind*III hoặc *Sma*I, rồi sau đó cả hai tay cho tiêu hóa với *Bam*HI. Có một đoạn lớn và một đoạn nhỏ được phân lập. Một đoạn chứa một cos site cộng với nguồn tự tái bản DNA (*Ori*), và một marker có tính chọn lọc với kháng sinh (ampicillin). Đoạn còn lại chỉ có chứa một cos site. Hai đoạn này kết hợp lại với nhau, thêm vào đó là genomic DNA với kích thước chọn lọc, có đầu cuối thích ứng (end) (*Bam*HI, *Sau*3AI, *Mbo*I), chúng khép lại với nhau thành vòng kín (packaged) vào trong phân tử lambda phage *in vitro*. Cuối cùng, DNA vòng kín này được sử dụng để chuyển vào tế bào *E. coli*. *E. coli* chứa cosmid DNA này, được chọn lọc bằng cách cho phát triển trên môi trường có chất kháng sinh (hình 3-7).

Kích thước DNA gắn vào vectơ				
Năm	1kb	10kb	100kb	1000kb
1970	Plasmid	Phage		
1980		Cosmid		YAC
			P1	
1990				BAC

Hình 3.7a: Tiến triển về khả năng mang các DNA có kích thước khác nhau của những vectơ

Bảng 3-2 : Số clone (N) cần để có xác suất 99% một clone đặc thù nào đó xuất hiện trong thư mục, với qui mô gắn vào là 40 kb, 150 kb, và 500 kb, trên vài loại cây trồng

Cây trồng	Kích thước genome (Mbp/1C)	Cosmid (40 kb)	BAC (150 kb)	YAC (500 kb)
Lúa	431	5.0×10^4	1.3×10^4	4.0×10^3
Lúa mì	15966	1.8×10^6	4.9×10^5	1.5×10^5
Bắp	2504	2.9×10^5	7.7×10^4	2.3×10^4
Mía	3000	3.5×10^5	9.2×10^5	2.8×10^4
Cà chua	953	1.1×10^5	2.9×10^4	8.8×10^3
Đậu nành	1115	1.3×10^5	3.4×10^4	1.0×10^4
Bông vải	2246	1.8×10^5	6.9×10^4	2.1×10^4

$N = \ln(1-P) / \ln(1-l/GS)$ trong đó N= số dòng, P=xác suất, l = kích thước gắn vào, và GS = genome size.

Hệ thống cosmid có nhiều ứng dụng khác nhau trong thực vật. Ứng dụng đầu tiên là công tác thanh lọc thư mục và lập bản đồ gen. Nhiều phòng thí nghiệm đã ứng dụng cosmid trong chuyển nạp gen thông qua *Agrobacterium* (Lazo và ctv. 1991, Meyerowitz 1992). Những vectơ này được cải tiến bằng cách thêm vào T-DNA border, những marker tương ứng, và đoạn gốc của DNA tự tái bản, có phổ khá rộng. Những vectơ này có thể được dùng để clone các DNA có kích thước lớn, rồi chuyển nạp trực tiếp vào cây trồng qua *Agrobacterium*. Thư mục cosmid đầy đủ của thực vật đã được tạo ra trên *E. coli* và *Agrobacterium*. Lazo và ctv. (1991) đã tìm thấy gen *E. coli lamdB* được thể hiện trên *Agrobacterium*, thể lambda infection receptor có thể được chuyển trực tiếp vào *Agrobacterium* mà không phải qua hoạt động con thoi của *E. coli* trước đó. Điều này cho phép sự thành lập thư mục của cosmid *Arabidopsis thaliana* 20.000 dòng, trong *Agrobacterium*.

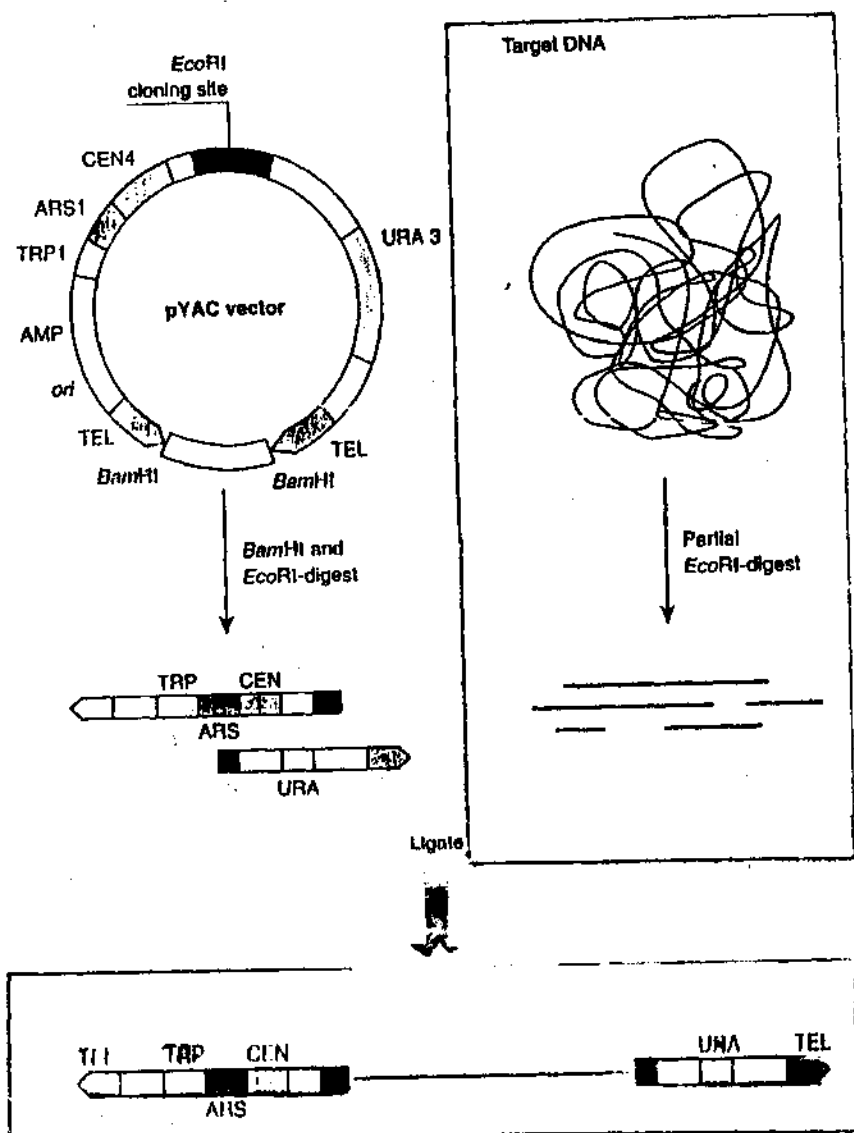
3.4. NHIỄM THỂ NHÂN TẠO LÀM TỪ MEN (YAC)

YAC vectơ đã được phát triển rất nhanh (Burke và ctv. 1987), đặc biệt trong kỹ thuật bản đồ gen, xây dựng contig, và kỹ thuật cloning các gen trên bản đồ. Vectơ YAC duy trì các DNA lạ trên men (yeast) như những nhiễm thể dây thẳng. Hệ thống vectơ này chứa đựng tất cả các phân tử của nhiễm thể thuộc nhóm eukaryotic (sinh vật bậc cao), bao gồm telomere, centromere và gốc của men thuộc DNA tái lập lại, cần thiết cho sự ổn định và duy trì nhiễm thể. Khả năng cloning của YAC ít bị hạn chế và chỉ tùy thuộc vào qui mô kích cỡ và chất lượng của DNA muốn cloning.

YAC viết tắt từ chữ yeast artificial chromosome, được ứng dụng rất sớm từ năm 1987. Nó có thể clone mọi DNA với bất cứ kích cỡ nào. Vectơ YAC sơ cấp (pYA4) và ký chủ của nó, men *S. cerevisiae* (AB1380 : MATa ura3 trp1 ade2-1 can1-100 lys2-1 his5) đã được sử dụng để làm ra thư mục thực vật (Burke và ctv. 1987). Hệ thống cơ bản của vectơ YAC bao gồm hai nhánh, cung cấp toàn bộ những yếu tố của một nhiễm thể sinh vật thượng đẳng: hai telomere và một centromere (*CEN4*), một đoạn gốc của DNA tự tái bản. Thêm vào đó, mỗi nhánh có một marker chọn lọc, nó có thể gây tác dụng phụ trong đột biến auxotrophic ở men chủ, đó là gen *TRP1* và *URA3*. (auxotrophic = dinh dưỡng thụ động). Cuối cùng, cloning site này được định vị trong gen tRNA suppressor, ký hiệu là *SUP4*. Khi *SUP4* tRNA được kết hợp, nó có thể ức chế đột biến *ade1*, mà đột biến này làm ra colony của men, có màu trắng. Nếu gen *SUP4* bị xáo trộn bởi gắn vào một đoạn DNA nào đó, thì đột biến *ade1* sẽ không còn bị ức chế nữa, và colony của men sẽ có màu đỏ. DNA được cloned trong vectơ YAC sẽ duy trì thành nhiễm thể của men có ít bản sao và phân ly một cách chính xác trong suốt giai đoạn gián phân giảm nhiễm và gián phân đẳng nhiễm (hình 3-8).

Kho lưu trữ YAC (library) được tạo ra cho nhiều loài thực vật như *Arabidopsis* (Grill & Somerville 1991), bắp (Edward và ctv. 1992), cà chua (Martin và ctv. 1992), lúa

(Umehara và ctv. 1995), v.v.. Vectơ chính được dùng là pYAC4 hoặc derivative của nó. Những vectơ khác cũng được xem xét là : pYAC-RC (rare cutter) có chứa nhiều vị trí cho restriction enzyme hiếm, dùng trong kỹ thuật cloning ; pYAC41 và pYAC45 có những promoter mang tính giải mã trên cả hai cloning site, người ta dùng nó trong kỹ thuật đẩy nhanh thể hệ của YAC end probes (chất thăm dò phía cuối của YAC) ; và vectơ pJS97, pJS98 được mô tả như một plasmid vai phải và vai trái, dùng trong việc phân lập cả hai đoạn cuối của YAC, trong kỹ thuật chromosome walking.



Hình 3.8 : Sự hình thành một nhiễm thể nhân tạo làm bằng men (YAC). Những vùng chủ yếu trong vectơ pYAC là : TEL-telomere của men (yeast); ARS1 chuỗi mã lập lại một cách độc lập; CEN4 - tam động của nhiễm thể số 4; URAA3 và TRP1 là những marker của men; Amp - kháng ampicillin thể xác định của pBR322; ori - nguồn gốc tự tái bản của pBR322

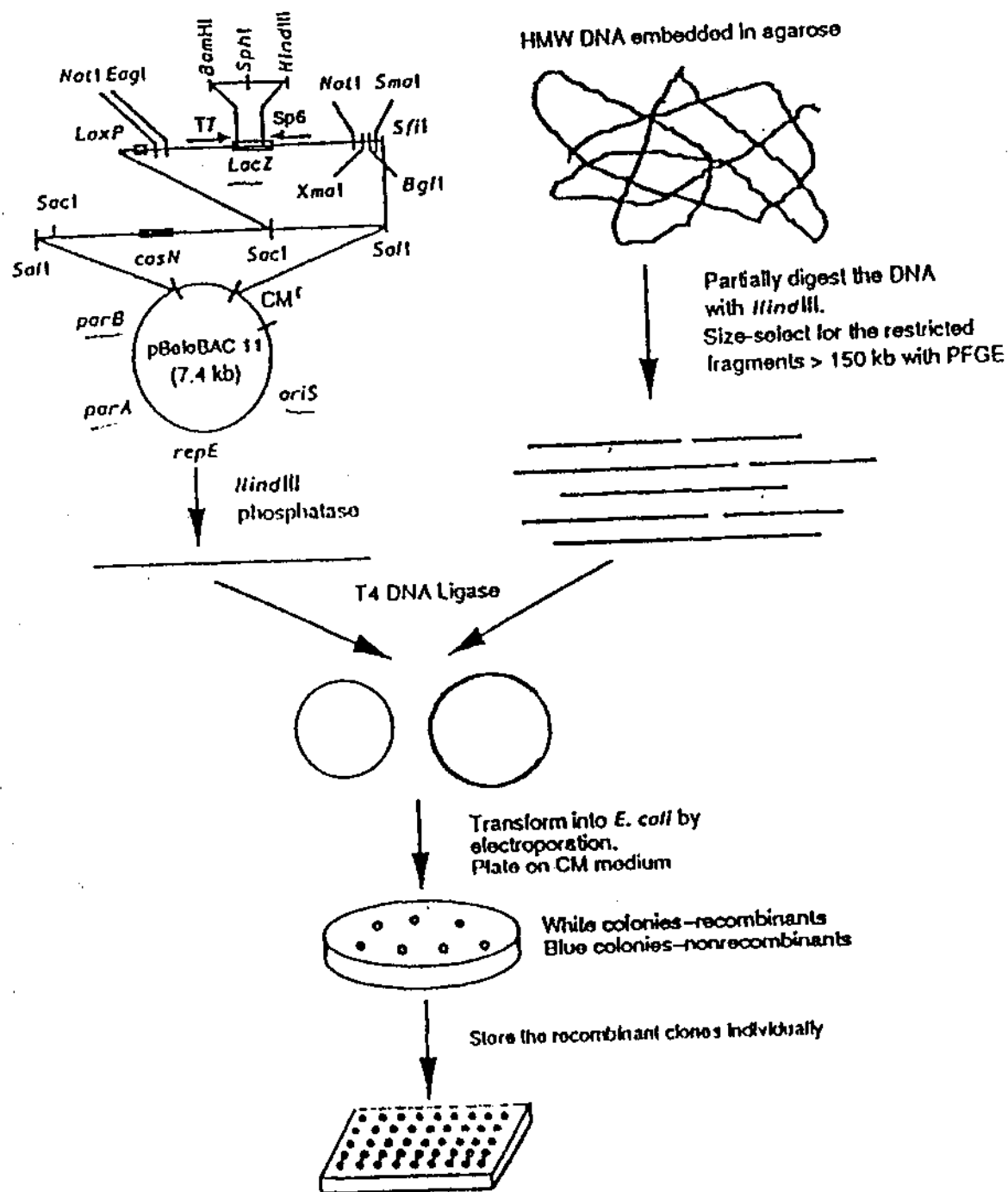
Trong kỹ thuật sử dụng pYAC4, người ta cho phân cắt YAC vectơ bằng *Bam*HI và *Eco*RI, rồi cho phản ứng dephosphoryl. DNA cũng được phân cắt từng phần với *Eco*RI, kích thước được chọn bằng hoặc gấp đôi, sau đó cho gắn vào vectơ này. Phản ứng kết gắn được chuyển vào thể cầu của *S. cerevisiae* theo phương pháp Burgess và Percivil (1987). Phương pháp chuyển nạp này là một trong những phương pháp tốt nhất, cho kết quả tần suất chuyển nạp $10^6 - 10^7$ colony unit trên microgram DNA vectơ kiểm chứng. Thể transformant được chọn lọc trên cơ sở uracil prototrophy và màu đỏ. Những dòng vô tính có màu đỏ được chuyển sang môi trường không có uracil và tryptophan để xác định sự có mặt của YAC trên cả hai nhánh. DNA của những recombinant clones được chuẩn bị bằng cách ấn tế bào của men vào trong agarose, sau đó tách khỏi thành tế bào ra bằng enzyme đặc biệt, rồi cho phản ứng phân hủy tế bào (lysis), xong rửa thật kỹ (Burke và ctv. 1987). DNA trong agarose được điện di trên điện trường CHEF (contour-clamped homogeneous electric field).

Trong trường hợp *A. thaliana* genome, library của cosmid gồm 17000 dòng vô tính được chia thành 800 contig. Những contig này đã được phân nhỏ ra bằng cách sử dụng YAC. Hwang và ctv. (1991) đã công bố YAC lai với 125 *A. thaliana* RFLP chiếm 1/3 genome này. Gần đây YAC được dùng để phân lập các gen thực vật bằng kỹ thuật cloning trên cơ sở bản đồ. Martin và ctv. (1993) đã tạo dòng DNA gen kháng bệnh trên cà chua *Pto*. Gen *Pto* có liên kết với những RFLP marker, được phân lập thành một bộ YAC phủ trên locus này. Trong trường hợp như vậy, người ta không cần kỹ thuật chromosome walking, mà cần kỹ thuật mới, với thuật ngữ "chromosome landing".

3.5. NHIỆM THỂ NHÂN TẠO LÀM TỪ VI KHUẨN (BAC)

BAC vectơ được phóng thích từ *E. coli* F factor plasmid, bao gồm các gen điều khiển số bản sao có tính giới hạn, và sự tự tái bản DNA có tính đẳng hướng (unidirectional). Cả hai tính chất này đều giúp cho plasmid được duy trì và ổn định. Khả năng cloning của BAC là 400 kb (Shizuya và ctv. 1992).

Shizuya và ctv. (1992) đã phát triển kỹ thuật BAC, viết tắt từ chữ "bacteria artificial chromosome". Oannou và ctv. (1994) đã phát triển kỹ thuật PAC (P1-derivative artificial chromosome), trong công tác cloning các DNA có kích thước lớn. Những vectơ này cho phép cloning các DNA có kích thước ít nhất 350kb trong *E. coli*. Cấu trúc căn bản của BAC/PAC vectơ được phát triển từ plasmid F nội sinh. Plasmid F rất cơ bản này bao gồm bốn vùng cần thiết, có chức năng ổn định plasmid và quyết định số bản sao. Đó là *ParA*, *ParB*, *OriS* và *RepE*. Cả hai *ParA* và *ParB* đều cần cho việc phân đoạn và sự ổn định plasmid. Thêm vào đó, *ParB* còn cần cho việc kết hợp với F factor khác. *OriS* là nguồn tự tái bản DNA, có tính chất đẳng hướng. *RepE* mang tín hiệu của protein E, cần trong quá trình tự tái bản từ *OriS*, và cần trong quá trình kiểm soát bản sao. Người ta cho vào gen



Hình 3.9a : Sơ đồ tổng quát của việc hình thành một kho lưu trữ (library) của nhiễm thể nhân tạo làm bằng vi khuẩn (BAC)

kháng chloramphenicol, dùng để chọn lọc các thể transformant có tính chất kháng sinh. Vectơ BAC nguyên thủy pBAC108L không tương hợp với gen *LacZ* để phục vụ cho việc chọn lọc nhờ phản ứng màu (Shizuya và ctv. 1992). Trong khi đó pBel AC11 có chứa gen *LacZ*, được dùng để xác định các dòng recombinant DNA. Vectơ PAC này thể hiện toàn bộ đặc điểm của hệ thống BAC (Oannou và ctv. 1994). Tuy nhiên vectơ PAC có chứa gen *SacB*, tạo ra sự chọn lọc dương tính đối với các dòng tái tổ hợp trong khi thành lập thư mục. *SacB* điều khiển sinh tổng hợp sucrose. Khi tế bào sinh trưởng trong môi trường saccharose, sự tổng hợp sucrose sẽ làm giảm saccharose ở mức độ gây độc cho *E. coli*. Cloning site của *Bam*HI nằm trong gen *SacB*, như vậy sự xáo trộn của gen này do gắn vào một DNA có kích thước lớn, cho phép sự sinh trưởng của tế bào thuận lợi trên môi trường saccharose. Vectơ có pUC19 link bao gồm : gốc tự tái bản DNA (*Ori*) với số bản sao rất cao, sẽ được dùng để phát triển các vectơ có lợi và loại thải sau đó trong quá trình chuẩn bị vectơ thành lập kho lưu trữ (library) (hình 3-9a).

Dòng *E. coli* được sử dụng phổ biến nhất cho kỹ thuật BAC cloning là DH10B (Hanahan và ctv. 1991). Đặc điểm chính của dòng này là có những đột biến ngăn trở (1) sự phân cắt DNA lạ do hệ men nội sinh (*hsd* / RMS) ; (2) sự phân cắt DNA có gốc methyl (5'-methylcytosine hoặc residue của methyladenine, 5'-hydromethylcytosine) (*mcrA*, *mcrB*, *mcrC* và *mrr*) và (3) sự tái tổ hợp (*recA*I).

Thí dụ pBel AC11: Trước tiên, vectơ được tiêu hóa với *Hind*III, rồi dephosphoryl hóa để ngăn ngừa hiện tượng tự dính lại. Kế đến, DNA trọng lượng cao phân tử (HMW DNA) cũng được tiêu hóa từng phần với *Hind*III. Sau đó DNA có kích thước > 150kb sẽ được lọc ra trên CHEF gel. Woo và ctv. (1994) nhận thấy rằng, người ta cần phải thực hiện chọn lọc cả hai kích thước DNA để gia tăng kích thước trung bình của BAC, nơi gắn DNA vào, bằng cách làm bẫy bắt các DNA có kích thước nhỏ sau giai đoạn chọn lọc kích thước đầu tiên. Cuối cùng, vectơ và DNA được kết dính lại, và dùng kỹ thuật mở lỗ bằng điện (electroporation), cho chúng xuyên qua *E. coli*. Các thể transformant tái tổ hợp được chọn lọc trên môi trường có chứa chloramphenicol, 5-bromo-4-chloro-3-indoyl- β -D-galactaside (IPTG). Khi người ta phát hiện ra các thể transformant tái tổ hợp xong, kích thước của nó sẽ được xét nghiệm bằng kỹ thuật có thuật ngữ là " simple DNA miniprepation". Sau đó, người ta tiêu hóa chúng bằng *Not*I để giải phóng các DNA dính trong vectơ, và điện di trên CHEF.

Vì hệ thống BAC và PAC còn quá mới mẻ, do vậy chỉ có vài loại thực vật là có library. Trên cao lương, có 13750 clone của BAC (Woo và ctv. 1994). Kho lưu trữ này ước khoảng 90% genome của cao lương, và 14% bị tạp với sequence của lục thể (chloroplast). Kho lưu trữ này được thanh lọc với 9 RFLP probe ngẫu nhiên, 6 của cao lương và 3 của bắp. Có ít nhất 1 BAC clone đã được tìm thấy đối với mỗi probe. Để đánh giá tính ổn định của clone, người ta dùng bốn BAC clone gắn giữa kích thước 280-315 kb, cho phát triển hơn 100 thế hệ và so sánh các vị trí phân cắt của enzym (restriction digestion). Người ta

không tìm ra được sự tái sắp xếp nào, chứng tỏ rằng DNA sequence của thực vật có thể được duy trì một cách ổn định trong BAC của *E. coli*. Cuối cùng, mức độ chimerism trong thư mục đã được thẩm định bằng phương pháp huỳnh quang *in situ* (FISH = fluorescence in situ hybridization). Mức độ chimerism của thư mục BAC trên cao lương rất thấp. Zhang và ctv. (1997) đã xây dựng thư mục BAC cho lúa thuộc nhóm japonica và indica. Kết quả ghi nhận tương tự như trên cao lương. Những kết quả đầu tiên này cho thấy BAC rất có triển vọng trong ứng dụng sinh học phân tử để phân tích genome thực vật.

3.5.1. Xây dựng kho lưu trữ BAC

Hiện nay, hệ thống BAC được phát triển như một công cụ có sức thuyết phục mạnh mẽ nhất trong công tác thành lập genome library (Shizuya và ctv.1992), vì những ưu điểm như sau:

- Nó có khả năng rất lớn gắn các đại phân tử DNA 100-300 kb
- Nó cho kết quả rất thấp về sự hình thành chimera (thể khảm)
- Nó có hiệu quả cao trong kỹ thuật cloning và phục hồi DNA
- Nó duy trì sự ổn định các DNA được gắn vào vectơ

Những clone có tính chồng lấp lên nhau được gọi với thuật ngữ "contig", có thể được phát sinh thông qua: (i) việc thanh lọc khuẩn lạc (colony), (ii) kỹ thuật chromosome walking, (iii) kỹ thuật DNA fingerprinting, (iv) hoặc STS marker (những marker được chuyển đổi từ RFLP) (viết tắt từ chữ sequence-tagged site).

Những nỗ lực đầu tiên có tính hợp tác quốc tế là xây dựng một thư mục BAC cho từng cây trồng, thí dụ như cây lúa. Hệ thống BAC có một vài thuận lợi hơn so với hệ thống YAC, như dễ xử lý *E. coli*, hiệu quả cao trong chuyển nạp, dễ phân lập các DNA gắn vào *E. coli*, dùng thể agarose gel thông thường để điện di, có tần suất thấp các chimeric clone (Shizuya và ctv. 1992, Woo và ctv. 1994, Wang và ctv. 1995). Do đó, hệ thống BAC có thể được xem có hiệu quả hơn hệ thống YAC, trong nghiên cứu xây dựng bản đồ vật lý, và phân lập tilling path thấp nhất (nghĩa là số dòng BAC chồng lấp thấp nhất, phủ trên chiều dài lớn nhất của genome cây lúa).

Muốn hợp nhất bản đồ vật lý với bản đồ di truyền, người ta phải dùng BAC clone trên bản đồ DNA marker. Những BAC clone này sẽ trở thành những ký hiệu (landmark) của bản đồ vật lý, nó đánh dấu số nhiễm thể và vị trí trên nhiễm thể. Những landmark này có thể được sáng tạo từ bản đồ STS trước đó, hoặc sự lai giữa các colony với RFLP marker. Nếu những marker này (STS hoặc RFLP) liên kết với những gen điều khiển tính trạng quan trọng về nông học, thì các landmarks đó sẽ cung cấp cho chúng ta những quan điểm đầu tiên về hiện tượng chromosome walking và kỹ thuật nhân bản vô tính các gen quan trọng.

Trong trường hợp cây lúa, người ta đã xây dựng kho lưu trữ BAC trên giống IR64 và Azucena (Yang 1997, Yang và ctv. 1997). Quy trình được thực hiện theo trình tự như sau:

Chuẩn bị DNA có phân tử cao (HMW) :

DNA có trọng lượng phân tử cao được ly trích từ mô cây lúa theo qui trình của Wang và ctv. (1995) sao cho tính chất cải tiến (modification) thấp nhất. Lấy 200 g lá lúa bốn tuần tuổi để ly trích DNA. DNA pellet được treo trong dung môi có 2,5ml SCE buffer và cho vào một lượng có thể tích tương đương agarose 2% có độ melting thấp. Sau khi xử lý proteinase K trong vòng 72 giờ ở 50°C, agarose này được thẩm tích (dialyse) hai lần khỏi TE buffer, cộng thêm PMSF (phenylmethyl sulfonylfluoride) trong vòng 1 giờ ở 50°C, rồi được điều tiết bởi *HindIII* buffer hai lần, ở điều kiện nhiệt độ trong phòng, 1 giờ. Sau khi lấy ra khỏi *HindIII* buffer, agarose này được xử lý ở nhiệt độ 65°C trong 10 phút cho phản ứng melting, và giữ ở nhiệt độ 37°C trong 5 phút. Phản ứng xong, cho thêm 1/10 thể tích 0,5 M EDTA. Điện di trên gel xung điện trường (PFGE) được tiến hành ngay sau đó. Đối với lần chọn thứ nhất, người ta điều chỉnh thời gian ramp từ 90-130 giây; 4,5 V/cm với góc quay 120° tạo xung (pulse), trong 18 giờ ở 14°C. Gel có chứa DNA với kích thước 200-600 kb được cắt và được đặt vào LMP agarose gel 1%, cho tiến trình chọn lọc lần hai. Lần này điều chỉnh thời gian ramp 5-10 giây, 4,5 V/cm với góc quay 120° tạo xung (pulse), trong 10 giờ ở 14°C. Sau khi PFGE, băng của DNA đậm đặc được cắt từ gel. Gel được thẩm tích hai lần khỏi TE buffer trong vòng 2 giờ. Kế đến cho nó ở điều kiện nhiệt độ 65°C, 10 phút để melting. Nó được giữ trong 5 phút ở 40°C. Người ta cho thêm 100 µl agarase, rồi ủ ấm 40°C trong 1,5 giờ. Sau 1 giờ tiêu hoá, 2 µl DNA được lấy ra để kiểm tra nồng độ, bằng cách điện di chúng trên agarose gel 0,6%.

Chuẩn bị phản ứng dephosphoryl của vectơ (dephosphorylation):

Các vectơ BAC, pBel AC 11 trong *E.coli*, DH10B được chuẩn bị.

50ml môi trường LB có chứa 30 µg/ml chloramphenicol được chủng vào từng colony của *E.coli*.

DNA của các vectơ được ly trích theo qui trình của Sambrook và ctv. (1989)

DNA của plasmid được thuần chủng bằng ly tâm theo gradient tỉ trọng của cesium chloride, cộng với sự hiện diện của ethidium bromide. Band của plasmid được lấy ra khỏi density gradient và ethidium bromide được loại ra nhờ ly trích nhiều lần với nước bão hòa isoamyl alcohol.

Mẫu DNA được hòa ra 5-fold với TE và ngưng kết với ethanol.

DNA của vectơ được dephosphorylated với HK phosphatase

Phản ứng kết gắn và chuyển nạp:

Đây là công đoạn cuối trong xây dựng thư mục BAC. Phản ứng kết gắn (ligation) được thực hiện trong 100µl thể tích, với tỉ lệ giữa DNA phân tử cao và DNA của vectơ đã được phản phosphoryl là 1:5.

Kết quả của giai đoạn này cho ra những thể recombinant có colony màu trắng.

Bảng 3-3: Ảnh hưởng của thời gian gây xung trên hiệu quả chuyển nạp, khả năng mang của các BAC clone (Yang 1997)

Thời gian gây xung ^a (milisec)	Voltage ^b (kv)	Colony ^c	Recombinant (%)	Khả năng mang (kp)	False-positive (%) ^d
5	1,25	45	97.8	93	0
10	1,25	125	87.2	109	0
15	1,25	137	86.9	118	0
20	1,25	129	89.9	104	0
25	1,25	59	67.8	95	2.5

a. Thời gian gây xung (pulse time) có nghĩa là Resistance x Capacitance

b. Lực của điện trường, KV/cm

c. Số colony / lần chuyển nạp dùng 1 µl trong cùng phản ứng kết dính

d. Phần trăm BAC clone không có màu trắng (không insert)

Mỗi BAC clone được ký hiệu theo số plate (X), số hàng (Y), và số cột (Z). Thí dụ BAC trong trường hợp IR64 và Azucena tại IRRI (Philippines) được theo dõi có 48 plate đánh số thứ tự từ P1 đến P48, có 16 hàng đánh số thứ tự theo mẫu tự từ A đến P, và có 24 cột đánh số thứ tự từ C1 đến C24. Muốn biết BAC có gắn được DNA của marker cần nghiên cứu khi lập bản đồ vật lý, nó phải được ghi nhận đầy đủ ở cả ba :plate, hàng và cột, khi điện di trên agarose gel, và dĩ nhiên trên DNA của IR64, Azucena.

Kho lưu trữ BAC trên lửa hiện nay có 18.432 clone trên IR64 (Yang 1997). Khả năng lớn nhất để nó mang thể insert là 364 kb. Có 75% BAC clone có khả năng mang 76-135 kb, trung bình là 107 kb

3.5.2. Thanh lọc kho lưu trữ bac với Rflp Marker

Nó bao gồm các bước chuẩn bị như sau :

- Lọc các phân tử có tỉ trọng cao
- Lai giữa các khuẩn lạc (colony)
- Phân lập các dòng chồng lấp lên nhau và hình thành contig
- Thực hiện chromosome walking

DNA của BAC được tiêu hóa với endonuclease trước khi đánh dấu. Clone dương tính được làm thuần nhờ lai và hợp nhất chúng trong contig, trên cơ sở DNA fingerprinting của những BAC clone dương tính.

Trên cây lúa, có rất nhiều gen kháng được công bố và được lập bản đồ liên kết gen ở nhiễm thể số 4 (Mohan và ctv. 1994, Wang và ctv. 1994, Yoshimura và ctv. 1995, Sebastian và ctv. 1996, Yang và ctv. 1997)

Để xác định những BAC clone này có liên kết chặt chẽ với những gen kháng đó, người ta tiến hành công việc chromosome walking từ 10 contig (bảng 3-4). Hai landmark là RZ740 và RZ675 được hợp nhất thành contig, có khoảng cách vật lý là 449 kb. Những BAC clone này liên kết chặt chẽ với gen kháng bệnh đạo ôn *Pi-5* (t) (Wang và ctv. 1994), gen kháng bạc lá *Xa-2* (Yoshimura và ctv. 1995). Ngoài ra, BAC clone được phân lập với RZ569 và RG214 liên kết chặt chẽ với gen kháng sâu năn (muỗi gây lá hành) *Gm2* và *Gm6* (Mohan và ctv. 1994). RZ262 phân lập thành BAC clone 14E16, liên kết chặt chẽ với gen kháng bệnh tungro và rầy xanh. Bảy BAC clone được phân lập từ RZ69 và chromosome walking hình thành một contig có qui mô 290 kb. Contig này liên kết chặt chẽ với gen kháng rầy nâu *Bph3* (Yang 1997).

Bảng 3-4 : BAC clone được phân lập bởi RFLP marker và kỹ thuật "chromosome walking" trên nhiễm thể số 4 của cây lúa.

RFLP marker	BAC clone	Khả năng gắn DNA vào vectơ (kb)
RZ905	6N21	105
	40N24	146
	*25M12	74
	48J9	156
	31A1	70
	11E3	95
	5 I18	128
RG329	8 O18	141
	*20 B6	89
	27 J1	67
	26 L5	67
	14 J6	77
	26 K8	70
	31 G20	101
RG214	21 I15	93
	24 E21	50

RZ269	24 D22	102
	38 J9	119
	*47 L10	128
	21 A19	130
	16 C1	104
	10 B13	142
	21 K19	90
	17 H4	97
	*29 M8	91
	7 P7	90
RZ467	24 K15	73
	42 J1	140
	34 K15	105
	32 P5	131
	25 H7	120
	15 O16	129
RG190	*6 N13	70
	*22 P6	37
RG454	27 O22	37
	26 I10	115
	26 H5	136
	29 C5	42
	43 P15	132
	34 M22	76
	23 J6	117
	39 K11	100
RZ169	*13 B12	124
	*23 C3	105
	31 M13	138
	43 M23	118
RG788	41 O22	113
	*2 E23	94
	40 O21	87

RZ569	6 M12	304
	20 D10	116
	*19 H3	79
	31 E20	86
	38 C2	136
	48 K24	121
	23 N23	85
RG449	8 G2	142
	*5 N8	111
	35 H11	104
	42 G7	115
	41 H7	133
RZ787	5 F1	104
RZ717	13 D8	61
RG864	25 B5	61
	3 I11	116
	15 J22	105
	48 O21	69
	30 A9	110
	22 N9	89
	41 L11	94
RZ262	14 E16	61
RZ23	45 F19	60
	48 F7	86
	31 P3	61
	26 N5	160
	21 G8	140
RZ889	38 H17	61
	41 F15	73
	4 N13	90

RZ740	4 L19	364
	43 G2	-
	22 G1	86
	21 F15	67
	24 A24	124
RZ675	43 G2	87
RZ565	4 F22	71
	16 D14	124
	35 G3	102
RZ879	34 M21	79
	*20 K21	101
RZ830	3 J16	73
RG143	*23 D10	104
	47 H7	120
RG163	23 K4	123
	39 O5	70
	38 B5	70
RG91	14 E5	67
RG396	13 C13	95
RG375	48 J8	147
RG908	1 C3	48
	*2 K9	146
	2 P24	121
RZ86	*36 C17	113
	1 B9	100
CDO680	47 O2	97
BCD135	*39 O20	101

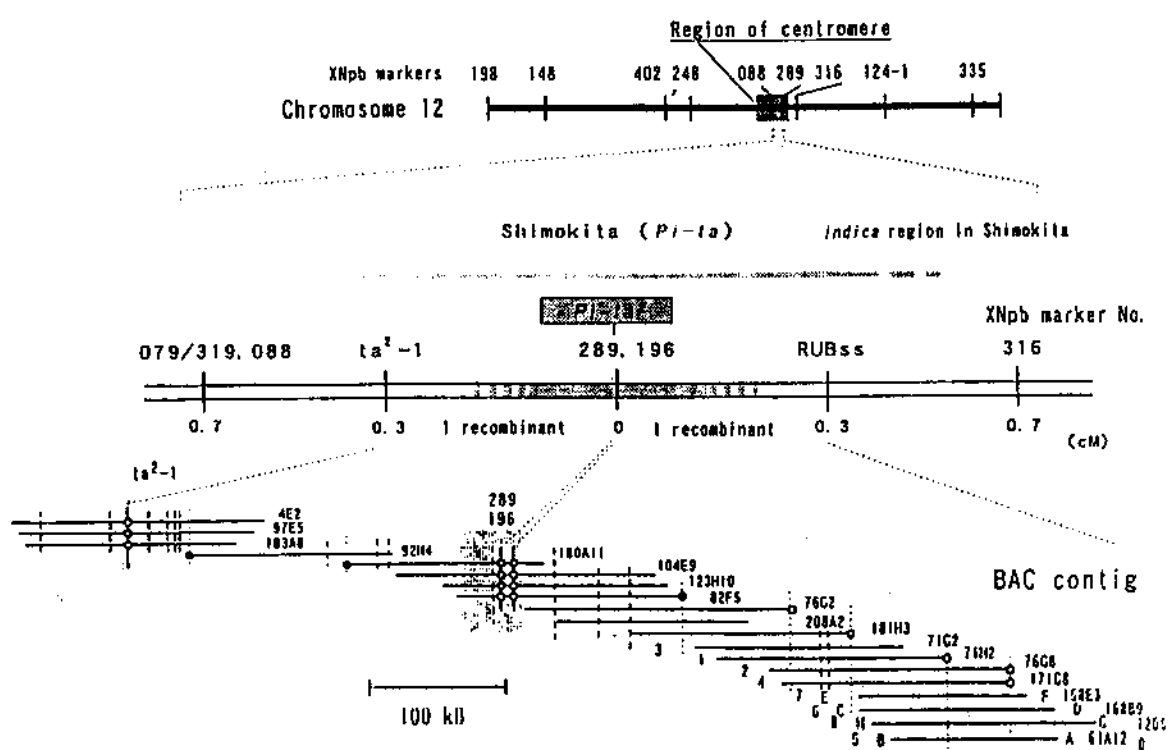
BAC clone có dấu hoa thị (*) được dùng làm probe

BAC clone in đậm được ghi nhận trong chromosome walking.

Hệ thống BAC có khả năng rất lớn để thành lập bản đồ vật lý có chất lượng cao, giúp ích cho nhà chọn giống khai thác hợp lý các vật liệu di truyền đã được đánh giá, nhờ phương tiện marker phân tử. Thí dụ trong trường hợp cây lúa, tại nhiễm thể số 4, có qui mô $3,75 \times 10^4$ kb

(tương đương size trung bình của tất cả các nhiễm thể), contig được hình thành từ BAC clone được phân lập với 28 RFLP marker sẽ cho chúng ta một độ bao phủ 15,1% nhiễm thể ($5,65 \times 10^3$ kb). Hơn nữa, contig này đã được thuần khiết bởi kỹ thuật Southern hybridization (sẽ thảo luận kỹ ở các phần sau), do đó nó rất đáng tin cậy. DNA có thể dễ dàng được phân lập trong BAC clone, cho nên mọi clone được nhặt ra từ kỹ thuật lai colony sẽ được kiểm tra bởi kỹ thuật DNA fingerprinting với sự tiêu hóa của *HindIII*. Kỹ thuật DNA fingerprinting cho chúng ta biết những phần nào của BAC có tính chất trùng lặp, phần nào không trùng lặp (non-overlap). Nhờ vậy người ta có thể hoàn thiện bản đồ vật lý có chất lượng cao, phục vụ cho công tác phân tích genome và chuyển nạp gen sau này.

Hình 3-9 là một ví dụ về contig trên nhiễm thể số 12 của cây lúa để nghiên cứu kỹ thuật cloning gen *Pi-ta²*, kiểm soát tính kháng bệnh đạo ôn (Nakamura và ctv. 1997)



Hình 3.9 : Hình thành một BAC contig đối với gen *Pi-ta²*, kiểm soát tính kháng bệnh đạo ôn cây lúa. Trên cùng là bản đồ RFLP ở nhiễm sắc thể số 12, tại tâm động. Ở giữa là Fine mapping nhờ RFLP và RAPD BAC. Vùng giới hạn của *Pi-ta²* và *Pi-ta* đã được làm dấu. Ở dưới cùng là sự chồng lấp của các BAC clone. Sự tương ứng giữa DNA marker(probe) với clone cộng tính được biểu thị bằng vạch chấm chấm (Nakamura và ctv. 1997).

TÀI LIỆU THAM KHẢO

1. Burgess PM, KJ Percivil. 1987. *Transformation of yeast spheroplasts without cell fusion*. Anal. Biochem. 163: 391-397
2. Burke DT, GF Carle, MV Olson. 1987. *Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors*. Science. 236:806-811
3. Edward KJ, H Thompson, D Edward. 1992. *Construction and characterization of yeast artificial chromosome library containing three haploid maize genome equivalents*. Plant Mol. Biol. 19: 299-308
4. Frischaud AM, H Lehrach, A Poustka, N Murray. 1983. *Lamda replacement vectors carrying polylinker sequences*. J. Mol. Biol. 170:827-842
5. Grill E, C Someville. 1991. *Construction and characterization of yeast artificial chromosome library of Arabidopsis which is suitable for chromosome walking*. Mol. Biol. 166: 557-580
6. Hwang I, T Kohchi, BM Hauge, HM Goodman. 1991. *Identification and map position of YAC clones comprising one-third of the Arabidopsis genome*. Plant J. 1: 367-374
7. Lazo GR, PA Stein, RA Ludwig. 1991. *A DNA transformation-competent Arabidopsis library in Agrobacterium*. Bio/Technology. 9:963-967
8. Martin G, S Brommonscheke, J Chungwongse. 1993. *Map-based cloning of a protein kinase gene conferring disease resistance in tomato*. Science. 262: 1432-1463
9. Martin GB, MW Ganai, SD Tanksley. 1992. *Construction of yeast chromosome library of tomato and identification of cloned segments linked two disease resistance loci*. Mol. Gen. Genet. 235: 25-32
10. Meyerowitz E. 1992. *Vectors for plant transformation and cosmid libraries*. Gene. 117:161-167
11. Mohan M, S Nair, JS Bentur, UP Rao, J Bennett. 1994. *RFLP and RADP mapping of the rice Gm2 gene that confers resistance to biotype 1 of gall midge (Orselia oryzae)*. Theor. Appl. Genet. 87:782-788
12. Nakamura S, S Asakawa, N Ohmido, K Fukui. 1997. *Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene Pi-ta² using a high representative rice BAC library*. Mol. Gen. Genet. 254:611-620
13. Oannou PA, CT Amemiya, J Grames. 1994. *A new bacteriophage P1-derived vector for the propagation of large human DNA fragments*. Nature Genet. 6:84-90

14. Sambrook J, EF Fritsch, T Maniatis. 1989. *Molecular cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press. New York. 1: 25-26
15. Sebastian LS, R Ikeda, N Huang, T Imbe, WR Coffman, SR McCouch. 1996. *Molecular mapping of resistance to rice tungro spherical and green leafhopper*. *Phytopathology*. 86: 25-30
16. Umehara Y, A Inagaki, H Tanoue, Y Yashukochi, Y Nagsamura, S Saji, Y Otsuki, T Fujumura, N Kurata, Y Min e. 1995. *Construction and characterization of a rice YAC library for physical mapping*. *Mol. Breed*. 1:79-89
17. Wang GL, DJ Mackill, M Bonman, SR McCouch, M Champoux, R Nelson. 1994. *RFLP mapping of genes conferring complete and partial resistance to blast in a durably resistant rice cultivar*. *Genetics*. 136: 1421-1434
18. Wang GL, TE Holsten, WY Song, HP Wang, PC Ronald. 1995. *Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus*. *Plant J*. 7: 525-533
19. Woo SS, J Jiang, BS Gill, AH Paterson, RA Wing. 1994. *Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor*. *Nucleic Acids Res*. 22: 4922-4931.
20. Yang D. 1997. *Construction of a bacterial artificial chromosome library and initial physical map of rice genome*. Ph.D Thesis. Wuhan University, China.
21. Yang D, A Parco, S Nandi, P Subudhi, Y Zhu, G Wang, N Huang. 1997. *Construction a bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4-specific RFLP markers in rice*. *Theor. Appl. Genet*. 95:1147-1154
22. Yoshimura S, A Yoshimura, N Iwata, SR McCouch, ML Abenes, MR Baraoidan. TW Mew, RJ Nelson. 1995. *Tagging and combining bacterial resistance genes in rice using RAPD and RFLP markers*. *Molecular Breeding*. 1: 375-387

Chương IV

LIÊN KẾT GIỮA GEN VÀ MARKER

DNA marker là những marker phân tử được tạo ra nhờ kỹ thuật phân cắt DNA bằng những restriction endonuclease. Khi sáng tạo ra những đoạn DNA mới, người ta đã cố gắng tìm kiếm một sự liên kết giữa marker và gen định vị trên nhiễm thể nào đó. Điều này hoàn toàn có thể thực hiện được nhờ công trình có tính lịch sử của Morgan về liên kết gen và khoảng cách di truyền (tính bằng centi Morgan, viết tắt là cM). Bản đồ di truyền đơn giản của ruồi giấm đã được phát hiện rất sớm vào năm 1925. Những tính trạng có biểu hiện giống nhau về di truyền được xếp vào cùng một nhóm liên kết gen, trên cùng một nhiễm thể. Những tính trạng này được xác định trên nhiễm thể tùy thuộc mức độ liên kết của nó. Sự sắp xếp một cách độc lập các nhiễm thể giúp cho việc định vị của loci trong các nhóm liên kết gen. Sự xuất hiện của hiện tượng quần chéo (crossing over) trong gián phân giảm nhiễm giúp cho việc xác định thứ tự của các loci trong nhiễm thể. Khoảng cách di truyền trên bản đồ được đo bằng tần suất tái tổ hợp gen (recombination). Gen thể hiện bản chất di truyền sẽ được liên kết với một tính trạng hình thái nào đó mà người ta có thể đo đếm được - gen đó có thể xem như là marker gen. Thí dụ liên kết gen giữa lá mầm có màu tím và tính kháng rầy nâu của một vài giống lúa ở Đông bắc Ấn Độ. Khi giống kháng có lá mầm màu tím được lai với giống nhiễm có lá mầm màu xanh, sẽ có 90% cơ hội để con lai F₂ có màu tím, thể hiện tính kháng rầy nâu. Như vậy lá mầm màu tím được xem như marker đối với tính kháng rầy nâu.

Việc lập bản đồ gen và chọn lọc nhờ marker hình thái như vậy sẽ rất chậm. Số marker quá ít và chỉ có ở qui mô hình thái (cơ quan).

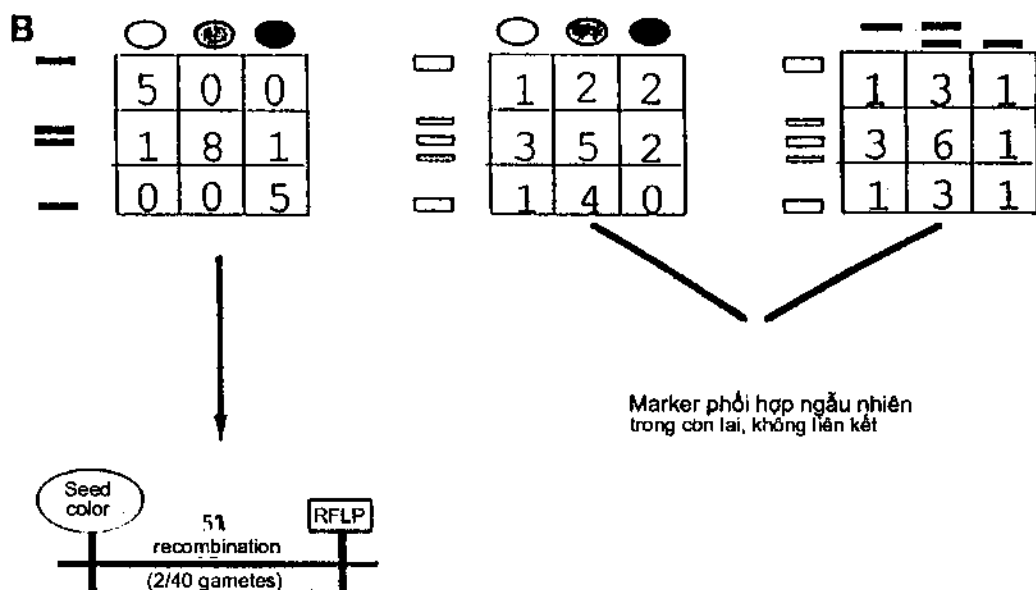
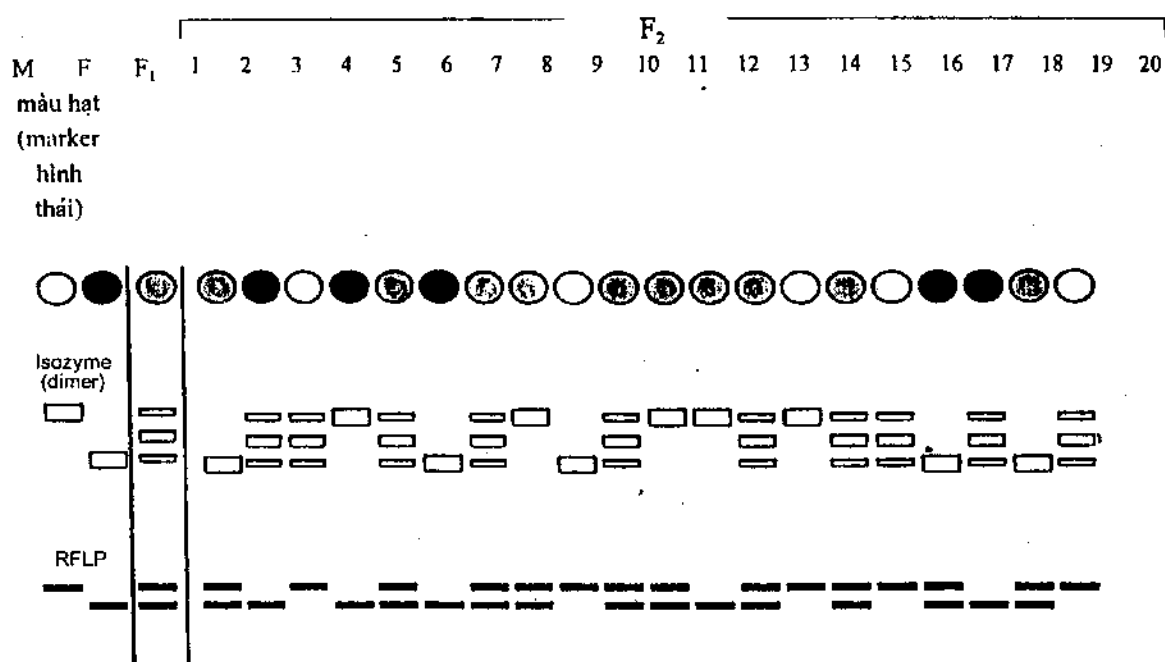
Việc áp dụng isozyme marker đã làm thay đổi theo chiều hướng tốt hơn, nhưng số marker cũng quá ít, không thỏa mãn cho nhu cầu nghiên cứu.

DNA marker đã được chứng minh có tầm quan trọng hơn về lâu dài, do số lượng của nó lớn hơn rất nhiều lần isozyme marker (Tanksley và ctv. 1980). Việc áp dụng DNA marker dễ dàng hơn và tương lai sẽ rẻ tiền hơn nhờ sự cải tiến không ngừng của các nhà khoa học (hình 4-1).

4.1. CÁC LOẠI DNA MARKER

Về căn bản, bất cứ chuỗi mã DNA nào được dùng để phân biệt giữa hai cá thể, hai dòng hoặc giống khác nhau, đều có thể được xem như là một DNA marker. Các DNA marker có thể được chia thành hai nhóm như sau :

- PCR-based : ALP, AFLP, SSR, SSCP
- DNA / DNA hybridization-based : RFLP, minisatellite



Hình 4.1 : M : giống mẹ, F : giống bố. Sau khi lai, con lai F₁ cho tự thụ để sản xuất quần thể F₂. So sánh marker hình thái, isozyme marker và RFLP marker, đối với màu vỏ hạt đậu

Các marker thuộc nhóm PCR-based có thể được chia nhỏ thành :

MAAP

marker ngắn, có tính ngẫu nhiên : RAPD, AP-PCR

DAF, AFLP

và amplicon đơn: ALP, SSR, SSCP

Những lợi ích của DNA marker so với marker hình thái và isozyme marker là

- đo lường trực tiếp các vật liệu di truyền
- có nhiều marker trong quần thể
- đo lường không chỉ phối ảnh hưởng môi trường và ảnh hưởng có tính chất phát triển

Bảng 4-1 : Các loại DNA marker

Marker	Tên đầy đủ
RFLP	Restriction fragment length polymorphism
ALP	Amplicon length polymorphism
AFLP	Amplified fragment length polymorphism
RAPD	Random amplified polymorphic DNA
DAF	DNA amplification fingerprinting
SSR	Simple sequence repeat (microsatellite)
AP-PCR	Arbitrary primer-PCR
SSCP	Single strand conformation polymorphism
MRDHV-DNA	Moderately repeated, dispersed, and highly variable DNA (minisatellite)

4.2. RFLP MARKER

Trong các DNA marker, RFLP là marker được dùng phổ biến nhất và được biết nhiều nhất. Trong nhiều trường hợp, thuật ngữ RFLP marker đồng nghĩa với DNA marker. Có nhiều giai đoạn để tạo ra chất thăm dò RFLP marker và tạo ra thể đa hình DNA.

4.2.1. Phân lập DNA

Trước hết chúng ta phải chuẩn bị chất thăm dò của marker và bộ lọc DNA lai. Đó là giai đoạn có tính chất quyết định trong phân tích RFLP. Thí dụ chúng ta cần có DNA từ

mô lúa với số lượng nhiều và chất lượng tốt. Trong quá trình phân lập DNA, người ta phải lấy mô lúa từ trong nhà kính hoặc trên đồng ruộng, mô này được nghiền bằng cối trong nitrogen lỏng. Bởi vì thành tế bào cây lúa rất cứng và rất khó phá vỡ. Nitrogen lỏng làm cho các mô trở nên lạnh và dẻo, như vậy mô có thể được nghiền mịn. Chất đệm khi ly trích DNA có chứa SDS, chất này có thể phân giải màng tế bào và màng nhân, phóng thích DNA vào dung dịch. SDS là chất tẩy khá mạnh, nó có thể làm biến đổi protein ở nhiệt độ 65°C. Sự biến đổi này làm cho các DNA gắn với protein bị tách rời ra khỏi DNA cần nghiên cứu. Protein và chất cặn của tế bào được tách rời ra khỏi DNA bằng phenol / chloroform hay bằng phương pháp kết tủa chọn lọc của những chất cặn tế bào với potassium acetate. DNA ở pha lỏng sẽ được kết tủa bằng cách thêm vào isopropanol.

Phân lập DNA bằng phương pháp nói trên chỉ áp dụng với DNA có kích thước lớn hơn 30 kb có một ít tạp chất như polysaccharides. Số lượng nhỏ của tạp chất như vậy có thể chưa gây ra vấn đề gì khi phân tích Southern. Nhưng chúng sẽ gây khó khăn khi chạy điện di trên gel, do đó chúng ta phải hết sức thận trọng trong lúc phân lập DNA.

Nguồn gốc của mô cũng ảnh hưởng đến số lượng và chất lượng DNA. Mô của cây khỏe và trẻ cho kết quả phân lập DNA tốt nhất. DNA được phân lập trên mô già hoặc trên loài lúa hoang thường bị tạp do polysaccharides và cho kết quả rất thấp.

DNA được phân lập có thể tồn trữ trong điều kiện -20°C. Nhưng chúng ta phải lưu ý rằng, một chút thay đổi nhỏ nào trong quá trình đông lạnh cũng có thể làm phân hủy DNA, tránh đông lạnh nhiều lần, hoặc làm tan băng nhiều lần DNA.

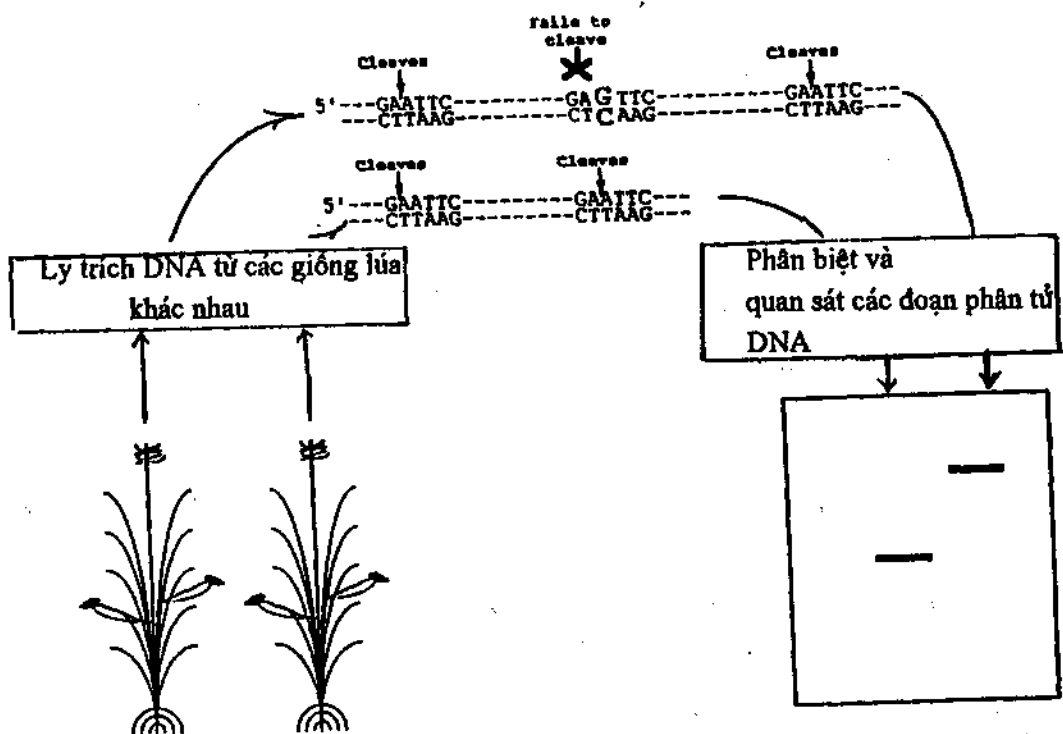
Trong quá trình phân lập, người ta phải sử dụng nhiều dung môi. Có hai dung môi hết sức quan trọng đó là EDTA và Tris.

Tris (Trisma, base) đã được sử dụng với khả năng đệm rất có hiệu quả trong dung môi. Ở pH 8.0, DNA rất ổn định. Trong điều kiện môi trường acid, purine rất dễ bị loại thải, tiến trình này được gọi là "depurination". Cầu nối phosphodiester dễ bị phá vỡ, tạo ra kết quả depurination. Tuy nhiên tính chất này vẫn được sử dụng trong Southern blotting. Do đó, người ta phải kiểm soát pH của dung môi ở mức 8.0 khi dùng Tris.

DNA có thể bị hỏng bởi các enzyme làm cho các mẫu DNA bị tạp. Chỉ cần chất gây tạp ở thể vệt cũng đủ để gây cho DNA bị hỏng. Enzyme này có thể truyền từ tay của người làm thí nghiệm, hoặc từ không khí, do vậy chúng ta phải rất thận trọng. Để ngăn ngừa khả năng DNA bị thủy phân bởi enzyme, người ta phải sử dụng thêm EDTA. Tất cả các enzyme thủy phân DNA được biết hiện nay đều cần Mg^{++} làm cofactor. EDTA có thể kềm giữ Mg^{++} và làm bất hoạt nó trong dung môi. Không có Mg^{++} , enzyme không thể hoạt động.

4.2.2. Sự phân cắt DNA

Có một nhóm enzyme thủy phân DNA được ghi nhận chuyên tính đối với một chuỗi mã DNA nào đó. Nhóm enzyme này được gọi với thuật ngữ là "restriction endonuclease". Tiến trình thủy phân DNA bằng restriction endonuclease được gọi với thuật ngữ "restriction digestion" [sự tiêu hóa tại điểm xung yếu, có tính chất hạn chế]. Để hiểu rõ tại sao người ta phải dùng từ "restriction" trong trường hợp này, chúng ta hãy xem lại khái niệm có tính chất giáo trình về di truyền thể phage. Người ta đã tìm ra rất nhiều restriction endonuclease, với hơn 100 enzyme có giá trị kinh tế, mỗi enzyme này có tác dụng chuyên tính với một chuỗi mã di truyền nào đó. Trong hầu hết các trường hợp, đoạn mã di truyền này có tính chất tự tái bản, được cắt theo dạng palindromic (5'-3'). Thí dụ, endonuclease *EcoRI* chỉ phản ứng trên chuỗi DNA tại G/AATC. Theo tính chất của cấu trúc palindromic tại restriction site, người ta có thể viết một nửa chuỗi mã di truyền DNA, nếu một nửa trước đó, kể từ restriction site đã được biết rồi (hình 4-2).



Hình 4.2 : Phát hiện DNA marker bằng phương pháp sử dụng RFLP marker. DNA được ly trích từ hai hoặc nhiều cá thể khác nhau, rồi phân cắt chúng bằng enzyme phân cắt hạn chế, trên chuỗi mã xác định đặc biệt của nó (thí dụ *EcoRI* là 5'-GAATTC). Hai cá thể sẽ thể hiện khác nhau về chuỗi mã DNA tại mỗi chuỗi mã xác định. Enzyme sẽ cắt chỉ tại một điểm, không cắt ở chỗ khác (dấu X), phát sinh ra nhiều đoạn phân tử có thuật ngữ là "restriction fragment". Nhờ điện di trên gel, những đoạn phân tử này được phân biệt do chiều dài khác nhau của nó. Đoạn ngắn chạy nhanh hơn đoạn dài hướng về cực dương. Nó được quan sát bằng mắt thường nhờ kỹ thuật chụp trên phim bức xạ với DNA thẩm dò có đánh dấu phóng xạ.

Nhiều đoạn mã được ghi nhận gồm có 6 cặp base [bp] của DNA. Trong một đoạn mã ngẫu nhiên nào đó, tần suất của vị trí này sẽ là 4096 bp (4^6). Nhóm enzyme ghi nhận 6 bp DNA được gọi là "6 bp cutter" trong phòng thí nghiệm. Nhóm enzyme khác ghi nhận 4 bp DNA, thí dụ như *Alu* I sẽ tiêu hóa chuỗi DNA có mang mã AGCT, được gọi là "4 bp cutter"

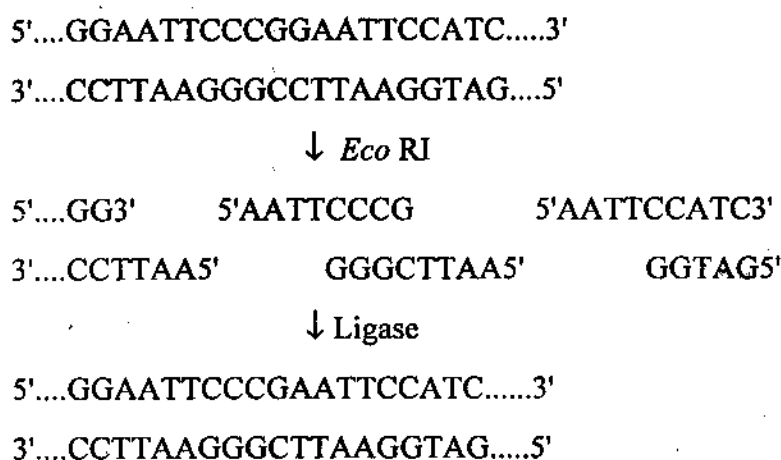
DNA sẽ được đưa vào một bộ sưu tập DNA thống nhất sau khi hoàn tất việc phân cắt các chuỗi mã thành những fragment riêng biệt.

Tất cả những fragment này sẽ có những phần cuối rất đặc biệt. Các fragment sau khi được tiêu hóa bằng *Eco* I, mang vị trí 5' ở điểm cuối - còn gọi là "sticky end". Các sticky end này có thể tác động qua lại với nhau. Nếu điều kiện cho phép, thí dụ như sự có mặt của Ligase và ATP, những sticky end kết hợp với nhau theo dạng đồng hóa trị, trước khi tiêu hóa.

Hiện nay, khi chúng ta mua enzyme, chúng ta phải lưu ý chất đệm của nó, với nồng độ x10. Các enzyme khác nhau yêu cầu những điều kiện khác nhau. Thông thường, phản ứng tiêu hóa xảy ra với điều kiện < 30% (vol) DNA, < 10% (vol) enzyme, 10% chất đệm của enzyme, < 50% nước. Nếu >30% (vol) DNA, > 10% (vol) enzyme, phản ứng tiêu hóa sẽ không thể xảy ra.

4.2.3. Điện di trên gel

Độ lớn của các fragment sau khi tiêu hóa biến thiên trong khoảng 30 kb, có khi nhỏ hơn 100 bp. Để xác định các DNA fragment thông qua kỹ thuật phân tích Southern, người ta điện di các DNA này trên agarose gel.



Hình 4.3b: Sự cắt đoạn do *Eco* RI và sự nối lại do ligase

Agarose là một trong các dạng của polysaccharide (hình 4-3). Theo tính chất của nó, các nhà sinh học phân tử đã khai thác đặc tính thể gel trong điện di (hình 4.3) để phân loại các DNA. Agarose sẽ tạo thành hạt agarose sau khi tan (melting) ở nhiệt độ cao, hoặc đun

sôi trong vài phút. Khi nguội lại, những hạt agarose sẽ kết tụ lại với nhau [gelling]. Giữa những hạt như vậy, có những lỗ rất nhỏ. Kích thước của những lỗ này có thể xê dịch chút ít tùy theo nồng độ của agarose gel. Trong điện trường, DNA di chuyển từ điện cực âm sang điện cực dương, vì DNA mang điện âm (do tính chất của gốc phosphate). Khi DNA di chuyển qua các lỗ của agarose, sự cọ xát giữa hạt agarose và phân tử DNA tạo ra lực kháng làm ngăn cản sự chuyển dịch này của DNA. DNA có phân tử càng lớn, thì lực cản càng mạnh. Do đó DNA có phân tử càng nhỏ, di chuyển càng nhanh. Nhờ vậy người ta phân loại được các DNA fragment trên agarose gel. Khi thay đổi nồng độ thể gel, kích thước lỗ giữa các hạt agarose cũng thay đổi. Nó sẽ trở nên lớn hơn khi nồng độ gel càng thấp. Khi kích thước lỗ to và phân tử DNA bé, thì sự khuếch tán sẽ gặp trở ngại. Người ta thấy rằng, cần có một sự tương xứng giữa hai yếu tố theo như bảng 4-2, để có được kết quả mong muốn trong khi nghiên cứu các đoạn DNA.

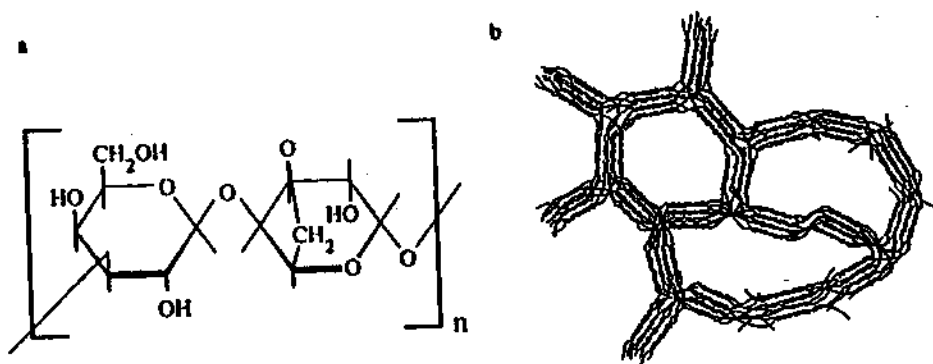
Bảng 4.2 : Sự tương ứng giữa nồng độ agarose và kích thước DNA fragment

Agarose (%)	Kích thước đoạn DNA [bp]
0.5	1 - 30
0.7	0.8 - 12
1.0	0.5 - 10
1.2	0.4 - 7
1.5	0.2 - 3
Acrylamide gel	< 500

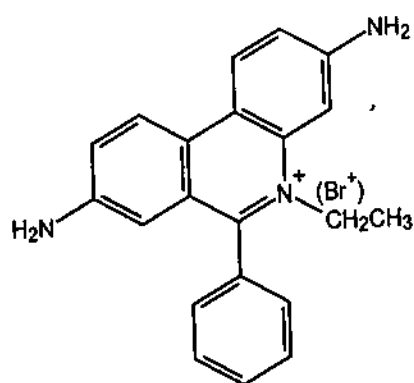
Sau khi điện di trên gel, các đoạn DNA được phân ra tùy theo trọng lượng phân tử. Người ta có thể quan sát chúng bằng mắt, nhờ kỹ thuật nhuộm màu DNA với ethidium bromide (hình 4-4) và thử nghiệm trong tia cực tím. Khi phân tích Southern, các đoạn DNA được chuyển vào các tấm lọc, nhờ lực mao dẫn. (hình 4.4).

4.2.4. Tạo dòng nhân bản vô tính DNA (clones)

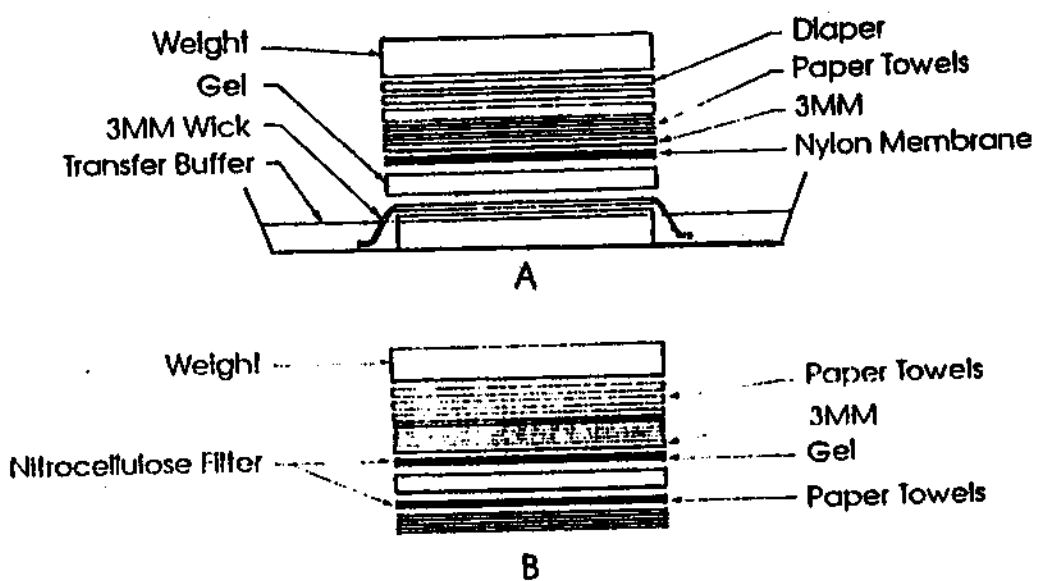
Có hàng triệu đoạn DNA sau khi cho phân cắt bằng *Pst* I chẳng hạn. Nhưng chúng không thể được sử dụng như RFLP marker, nếu từng đoạn này chưa được thanh lọc làm cho thuần khiết và chưa được khuếch đại (amplified) nhằm có đủ số lượng các đoạn DNA đồng nhất, thuần chủng. Kỹ thuật tạo dòng [cloning] các đoạn DNA nhờ plasmid, rồi khuếch đại plasmid trong một tế bào chủ của nó, như *E. coli* chẳng hạn, đã được áp dụng rất thành công.



Hình 4.3 : Cấu trúc của agarose (a), và hình dạng của nó trong một thể "gel" (b)



Hình 4.4 : Cấu trúc của ethidium bromide, phẩm nhuộm phân tử DNA



Hình 4.5 : Phương pháp Southern blot (sách hướng dẫn trong lab của Đại Học Cornell 1989)

4.2.5. Đánh dấu thăm dò và lai DNA

DNA đã bị phân cắt được chuyển vào tấm lọc. Các đoạn DNA này là một sự trộn lẫn của genome cây lúa. Muốn phát hiện ra chuỗi mã DNA đặc biệt nào đó, người ta phải áp dụng một kỹ thuật được gọi là lai DNA / DNA. Kỹ thuật này được khai thác căn cứ theo tính biến chất DNA [mở dây đôi thành dây đơn] và tính hoàn nguyên [nối lại dây xoắn đôi như cũ].

Các đoạn DNA sẽ được tách ra riêng nhờ kỹ thuật điện di trên gel. DNA này sẽ biến chất trước khi nó được chuyển qua màng lọc. Như vậy DNA trên màng lọc sẽ là DNA ở dạng dây đơn. Để xét nghiệm một DNA đặc biệt nào đó trên màng lọc, chúng ta cần có một DNA thăm dò [probe], nó giúp cho việc ghi nhận / hoặc lai với DNA đặc biệt nào đó, thông qua các nối base. Nếu DNA thăm dò đã được đánh dấu bằng chất đồng vị hay hóa chất nào đó, chúng ta sẽ xét nghiệm được DNA muốn biết, thông qua xét nghiệm chất thăm dò. Theo cách này, một chuỗi mã di truyền DNA cần nghiên cứu đã được xác định (hình 4-6).

4.2.6. RFLP marker

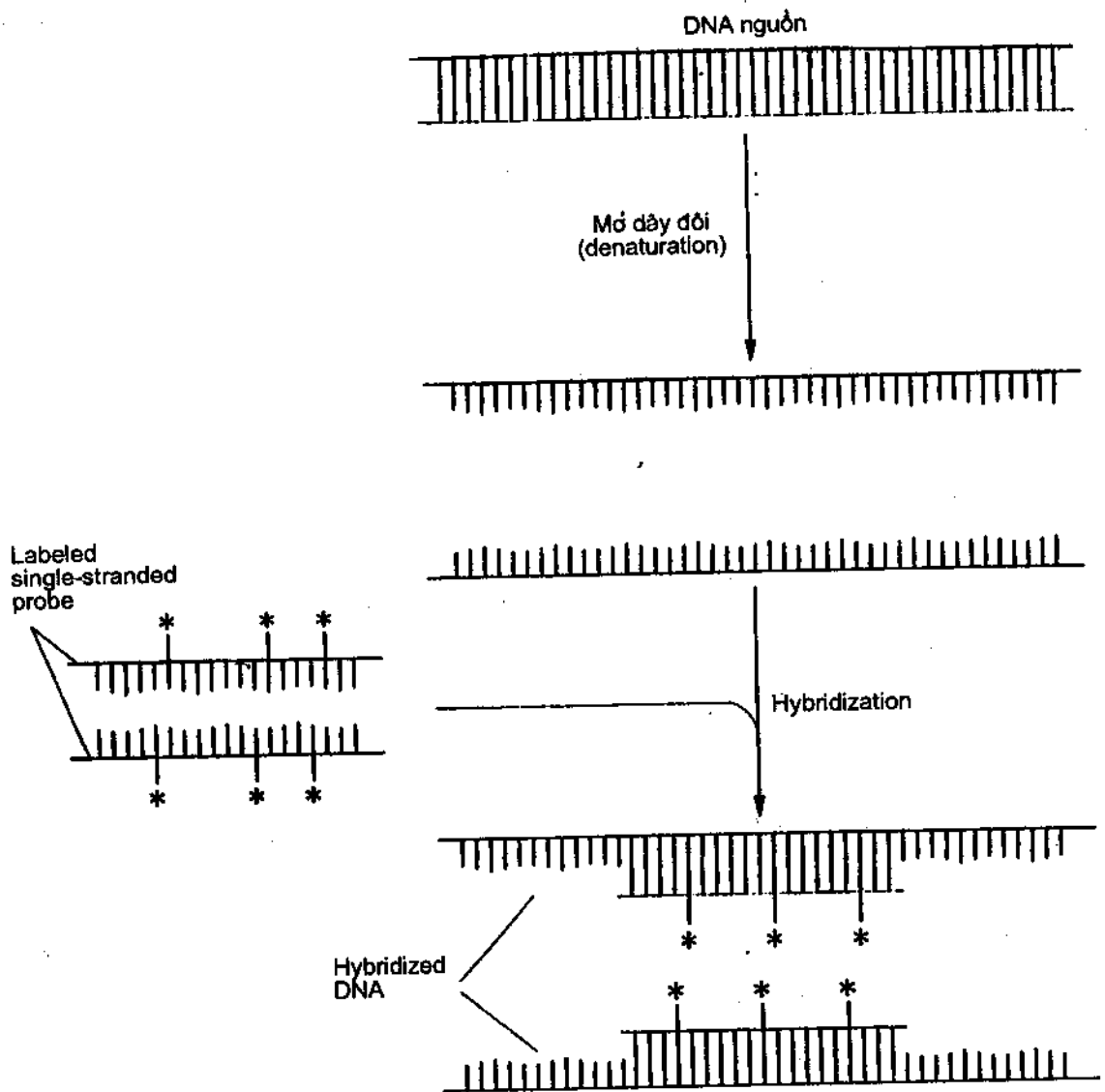
Đoạn DNA được xác định như phần mô tả ở trên được gọi là "restriction fragment". Chiều dài của nó được xác định bởi các vị trí phân cắt. Trong quá trình tiến hóa của cây lúa, sự thay đổi chuỗi mã di truyền DNA cũng có thể tạo ra sự thay đổi về chiều dài của đoạn DNA. Khi chúng ta so sánh hai giống lúa mà độ dài DNA của nó khác nhau, chúng ta có thể nói rằng, đó là thể đa hình về chiều dài của đoạn DNA, hay RFLP viết tắt từ chữ Restriction Fragment Length Polymorphism. Chất thăm dò được dùng để phát hiện ra thể đa hình được gọi là RFLP marker.

Theo ý nghĩa này, RFLP phải có tính chất như sau :

- Có số lượng cao trong mỗi quần thể
- Codominant (đồng tính trội)
- Đo trực tiếp DNA
- Không bị ảnh hưởng của môi trường
- Không bị ảnh hưởng do yếu tố có tính chất phát triển

4.2.7. Phương pháp Southern blotting

Các đoạn DNA bị cắt bằng restriction enzyme được cố định trong một màng đặc biệt mà người ta có thể dò xét nhiều lần bằng marker phân tử. Qui trình có thuật ngữ là "Southern blotting" (Southern là tên một nhà khoa học đề xuất phương pháp này).



Hình 4.6 : Lai DNA. DNA nguồn được phản ứng denature làm cho dây đôi được mở ra thành hai dây đơn, người ta gắn nó vào vật thể rắn như màng nylon hay màng nitrocellulose. DNA thăm dò (probe) được đánh dấu [100 - 1000 bp] bằng đồng vị phóng xạ. Nó cũng được phản ứng denature, rồi đưa vào cho lai với DNA nguồn. Sự lai do hiện tượng bắt cặp giữa các base tương ứng nhau của probe và DNA nguồn xảy ra. Màng lọc được rửa để loại bỏ DNA thăm dò chưa được lai. Người ta đem màng lọc này xét nghiệm, nếu có sự lai thực sự, nó sẽ được quan sát thể đánh dấu (tag) có phóng xạ. Dấu hoa thị trong hình là ""labeled tag" của probe.

Sự di chuyển thông qua mạch cực nhỏ của các đoạn DNA từ gel đến các loại màng lọc khác nhau - là một cuộc cách mạng trong nghiên cứu genome của sinh vật eukaryotic. Khả năng khám phá ra các chuỗi mã di truyền rất hiếm này, trong một quần thể phức tạp của những restriction fragments, đã mở đường cho việc tạo dòng các gen của eukaryotic. Việc sử dụng Southern để phân tích RFLP đã có một hiệu quả nổi bật trong chẩn đoán bệnh di truyền ở người. Việc cải tiến khái niệm cơ bản của Southern đã dẫn đến kỹ thuật lai DNA, hay kỹ thuật Northern cho phân tích RNA, và Western cho phân tích protein.

Các đoạn DNA tách riêng ra với nhau tùy theo kích thước của nó chạy trên agarose gel, và tình trạng mở dây đơn [denature]. Những đoạn DNA này được chuyển từ gel sang một thể rắn [tấm lọc nitrocellulose hoặc màng lọc bằng nylon], nơi nó bị cố định. Sau giai đoạn prehybridization (trước khi lai DNA), người ta khóa các vị trí trên màng - nơi quá trình lai DNA sẽ xảy ra, các DNA (gen liên kết với marker) sẽ gắn với nucleic acid thăm dò (probe, hay RFLP marker) được đánh dấu bằng phóng xạ. Quá trình lai giữa DNA (gen) và probe như vậy được gọi là "DNA hybridization". Sau khi hybridization, tấm lọc hay màng lọc được rửa để loại bỏ các probe không gắn, hoặc gắn yếu với DNA đang nghiên cứu. Tiếp sau đó, nó tự ghi trên biểu đồ phát xạ. Sau khi điện di, gel được chụp dưới tia cực tím. Các đoạn DNA xuất hiện thành các đốm liên tục. Nó sẽ cho tín hiệu khi thể hiện ra trên phim X-quang. Các sọc có tính đa hình (polymorphism) có thể được quan sát để đánh giá. Khi xử lý P^{32} phóng xạ, chúng ta nhớ phải tuân thủ những qui định an toàn trong phòng thí nghiệm. Hãy kiểm tra thường xuyên sự tạp nhiễm bằng Geiger counter (hình 4-5).

RFLP marker có khả năng sử dụng rất phong phú, nhưng qui trình thực hiện phức tạp, nguy hiểm đến sức khỏe người thực hiện, đắt tiền, yêu cầu DNA có số lượng và chất lượng rất cao. Do đó, người ta có xu hướng áp dụng những marker đơn giản hơn, an toàn hơn, trên cơ sở phản ứng chuỗi polymerase.

4.3. MARKER LÀ SẢN PHẨM CỦA PCR (PCR-BASED MARKER)

Phản ứng chuỗi polymerase được viết tắt PCR là tiến bộ kỹ thuật đã được ứng dụng rộng rãi trong những năm đầu của thập niên 1990. Kary Mullis người có công lớn trong phát hiện này, đã được lĩnh giải thưởng Nobel năm 1993.

4.3.1. Nguyên tắc cơ bản của PCR

Phản ứng chuỗi của polymerase thường được viết tắt là PCR (polymerase chain reaction). Đây là một kỹ thuật phân tử tạo dòng DNA rất đơn giản và hiệu quả. Thông qua PCR, hàng triệu đoạn DNA đồng nhất có thể được thu thập một cách dễ dàng từ một hỗn hợp các phân tử bao gồm RNA, protein, polysaccharide, DNA không có chức năng và DNA có chức năng di truyền. Người ta còn gọi nó là kỹ thuật tạo dòng DNA *in vitro*. Ngày nay, PCR được dùng rất phổ biến trong nhiều lĩnh vực thuộc về sinh học.

PCR là kỹ thuật xử lý *in vitro* các chuỗi mã hóa di truyền DNA bằng cách phát triển primer một cách đồng loạt trên các dây đơn DNA. Toàn bộ tiến trình được hoàn thiện do sự biến chất của DNA cần thiết, sự tác động của primer tại đầu của các dây đơn DNA này, kể đến là sự phát triển của các primer do phản ứng DNA polymerase (hình 4-7).

DNA polymerase là một enzyme có chức năng tổng hợp các dây đơn DNA. Trong điều kiện cho phép nào đó, tiến trình tổng hợp DNA này có thể được sao chép *in vitro*. Ứng dụng đầu tiên của sinh tổng hợp DNA *in vitro* gồm có tạo chuỗi mã DNA thông qua phương pháp gây ảnh hưởng đoạn cuối của dây dideoxy của Sanger, phương pháp giải mã DNA đánh dấu, v.v... Tất cả những ứng dụng này đều có trong giai đoạn 1 [chu kỳ 1] của sinh tổng hợp DNA.

Vào năm 1985, một nhóm các nhà khoa học của Cetus Corporation đã tuyên bố có một cách mới trong sinh tổng hợp DNA *in vitro*, với thành tựu bước đầu về phản ứng chuỗi polymerase có tính chất tự động. Họ đã sử dụng qui trình khuếch đại chuỗi mã di truyền beta globin :

- Tạo dây đơn DNA cần nghiên cứu
- Cho các primer tác động DNA cần nghiên cứu
- Thêm polymerase để tổng hợp dây đơn bổ sung
- Lập lại bước 1 đến bước 3 với nhiều chu kỳ

Kết quả cuối cùng là một đoạn mã hóa di truyền đặc biệt nào đó đã được khuếch đại lên rất nhiều lần. Sự khuếch đại này có thể được tính như sau :

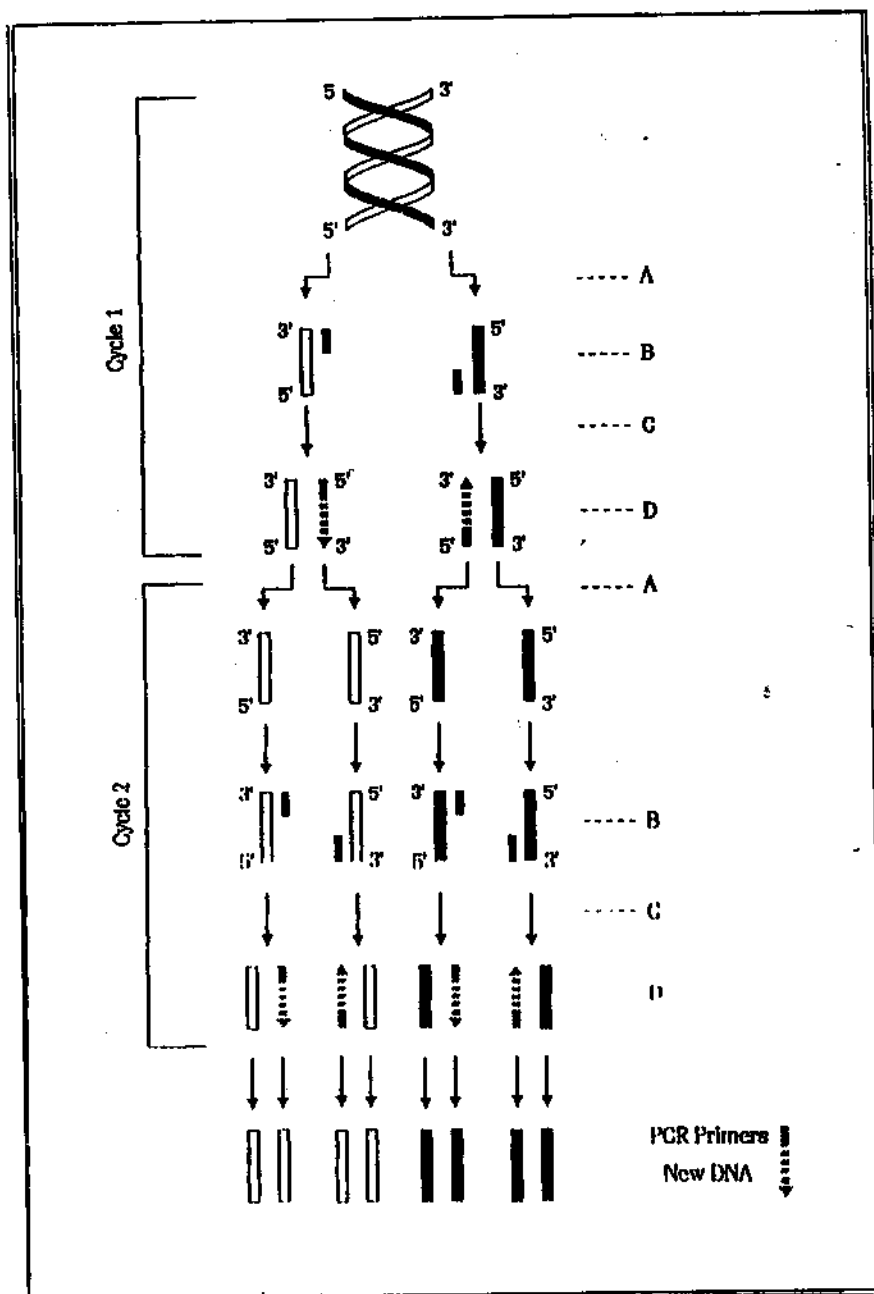
$$\text{Tổng số khuếch đại} = m \times 2^n$$

trong đó n là số chu kỳ và m là số copy của chuỗi mã hóa cần nghiên cứu. Giả sử chúng ta đạt hiệu quả là 100%, chúng ta sẽ có :

Số chu kỳ	Tổng số khuếch đại
1	2
5	32
10	1024
20	1048576
30	1.07×10^9

Có hai vấn đề chính trong nguyên tắc PCR :

1. Polymerase được sử dụng là những enzyme sống, nhạy cảm về độ nhiệt, phải được thêm vào trong mỗi chu kỳ.
2. Phải có một qui trình hướng dẫn cụ thể trong khi thực hiện từng giai đoạn của chu kỳ PCR.



Hình 4.7 : Phản ứng chuỗi polymerase (PCR). A : phản ứng mở dây đôi thành hai dây đơn nhờ nhiệt độ (denaturation). B : phản ứng của hai prime F và R tại hai đầu dây (annealing). C và D : phản ứng kéo dài phân tử DNA mới theo chiều dài của dây đơn DNA cũ (extension). Chu kỳ này lặp lại nhiều lần, làm cho phân tử DNA mục tiêu được khuếch đại lên nhiều lần (amplification)

Cả hai tiến trình nói trên đều nặng nhọc và bất lợi cho các nhà nghiên cứu, nếu họ có phương tiện quá hạn chế khi áp dụng PCR.

Tuy nhiên, trong thiên nhiên vẫn có những sinh vật có thể sống sót ở nhiệt độ cao. Thí dụ polymerase được phân lập từ *Thermus aquaticus* có khả năng là vật thể ổn định về nhiệt lượng. *Taq* polymerase có hai thuận lợi chính: [i] hồi phục sau khi tác động nhiệt không cần kéo dài, [ii] enzyme này rất hoạt động ở nhiệt độ cao, lúc đó việc tác động của các primer ở đầu dây đơn sẽ chuyên biệt hơn và sinh tổng hợp DNA sẽ nhanh hơn.

Sự phát minh ra máy thermocycler đã tạo điều kiện thuận lợi, giúp chúng ta có thể thay đổi nhiệt độ định kỳ theo một thời gian nhất định. Chiếc máy có tính chất chương trình hóa này đã hoàn thiện chu kỳ nhiệt một cách tự động, và người ta thường gọi nó là máy PCR.

DNA polymerase được dùng phổ biến là *Taq* polymerase, lấy từ *Thermus aquaticus*, có tính ổn định nhiệt rất cao. Một nửa chu kỳ của *Taq* polymerase được giữ ở nhiệt độ 94°C trong vòng 40 phút. *Taq* polymerase có một mức tối hảo về nhiệt độ cao trong sinh tổng hợp DNA [70-72°C]. Ở quãng nhiệt độ này, hoạt động đặc biệt của *Taq* polymerase có thể cao như 150 nucleotide/giây/enzyme. Do đó người ta phải cố gắng phát triển ở quãng nhiệt độ 70-72°C trong vòng một phút. Từ đó hàng kilo bp của các đoạn DNA có thể được tổng hợp.

Taq polymerase nhạy cảm với ion magnesium. Nồng độ tối hảo của Mg^{++} là 1,5 -2,5 mM. Vì deoxyribonucleotide triphosphate (dNTP) có thể gắn với Mg^{++} , cho nên nồng độ chính xác của magnesium tùy thuộc vào nồng độ của dNTP. Các thành phần khác ít mẫn cảm với *Taq* polymerase là KCl và Tris.

Chất đệm :

Tris (pH 8.4)	10mM
KCl	50mM
MgCl ₂	1.5-2.0 mM

Các thành phần khác như các chất tẩy không có tính ion, gelatin, NP40, và Tween 20 được thêm vào với số lượng nhỏ nhằm thúc đẩy hoạt động của *Taq* polymerase. Nồng độ hiện được dùng là 0.01% và người ta còn tiếp tục thử nghiệm.

Primer và dNTP :

Trong PCR tiêu chuẩn, người ta cần có một cặp primer. Sự khuếch đại chuyên biệt nào đối với một đoạn DNA cần nghiên cứu, đều phải tùy thuộc các primer tương xứng. Cả hai yếu tố : chiều dài primer và thành phần G+C đều có ảnh hưởng quan trọng đối với nhiệt độ trong quá trình tác động ở đầu dây đơn. Nhiệt độ này sẽ là $2x (\# \text{ của A+T}) + 4x(\# \text{ của G+C}) + 5^{\circ}\text{C}$. Nếu có 50% G+C, thì chiều dài của primer phải có kích thước tương ứng là 18-20 nucleotide.

Người ta mua các nucleotide ở các công ty hóa chất, có chất lượng tốt. Nếu nó được bảo quản ở dạng bột đông khô [freeze-dried powder]. Sau đó người ta phải điều chỉnh lại độ pH, trước khi sử dụng.

DNA mục tiêu (Target DNA) :

Người ta cần phải có các đoạn DNA mang mật mã đã biết trước, được gọi với thuật ngữ "target DNA" trong kỹ thuật PCR người ta sẽ nhận được những kết quả tốt nhất, nếu DNA thật sự thuần khiết, DNA từ mô lá. Số lượng DNA cần cho một phản ứng không nhiều lắm, chỉ cần 5-10 ng DNA thuần khiết, đủ để cho những kết quả theo mong muốn. Người ta cần có ethanol trong lần cuối của qui trình chuẩn bị DNA. Với 10% ethanol, nó vẫn chưa có thể gây ảnh hưởng đến hoạt động của *Taq* polymerase.

Điều ghi nhận quan trọng là : phải có một mM EDTA trong chất đệm TE trong khi sử dụng để hòa tan DNA. EDTA sẽ gắn với Mg^{++} , sao cho thể tích của DNA mục tiêu không vượt quá 1/5 của thể tích PCR

Phát hiện DNA có tính đa hình nhờ PCR :

Sản phẩm của PCR là những đoạn mã DNA. Khi DNA xuất phát từ hai dòng lai với nhau được khuếch đại lên nhờ các primer tại locus đặc biệt nào đó, thì trọng lượng phân tử của sản phẩm PCR này có thể rất khác nhau, vì có những thay đổi vật lý trên chuỗi mã DNA, nghĩa là, sự mất đoạn hay thêm đoạn DNA sẽ xảy ra trong vùng bị khuếch đại này. Thể đa hình được gọi là "amplicon length polymorphism" [ALP] phản ánh DNA có tính đa hình giữa các cá thể / các dòng lai / các giống lúa. Thuận lợi của ALP so với RFLP trong việc phát hiện này là : (1) nó rất nhanh, chỉ cần một ngày giúp ta tìm ra kết quả, (2) không có chất đồng vị, (3) rẻ tiền, (4) cần rất ít DNA. Bất lợi của ALP là người ta phải biết rõ chuỗi mã DNA đầu tiên khi tổng hợp primer.

4.3.2. RAPD marker

Một trong những giới hạn chính của PCR chuẩn đối với ALP marker là mọi thông tin về chuỗi mã di truyền đầu tiên phải được biết rõ trước khi chuẩn bị các primer tương ứng. Nhưng hiện nay, thông tin về các chuỗi mã này trên các vùng của genome cây lúa chưa được biết hết. Để phát hiện ra các thể đa hình DNA trên những vùng như vậy, người ta phải cải tiến kỹ thuật PCR. Vào năm 1990, có hai nhóm nghiên cứu đã thực hiện việc cải tiến này để phát hiện thể đa hình DNA bằng PCR tiêu chuẩn (William và ctv. 1990, Saiki và ctv. 1988, Micheltore và ctv. 1991)

Nguyên tắc : từ lâu người ta đã biết rằng việc khuếch đại một đoạn DNA đặc biệt nào đó, có thể thu nhận được kết quả thông qua điều kiện nghiêm ngặt PCR. Nhưng các sọc không có tính chuyên biệt gì vẫn thể hiện ra trong điện di, một khi sự nghiêm ngặt này bị lơ lửng. Đặc biệt trong trường hợp nhiệt độ ở giai đoạn tác động của primer tại đầu dây

đơn bị giảm thấp hơn T_m , sẽ có nhiều sọc bất thường thể hiện ra trên gel. Welsh và McClelland (1990) đã ghi nhận hiện tượng này và đã phát minh ra AP-PCR (arbitrary primer-PCR) (những primer có chuỗi mã ngắn, có tính ngẫu nhiên).

Trong AP-PCR, người ta sử dụng một primer đơn hay một cặp primer với một chuỗi mã điều hành (arbitrary) có khoảng 20 nucleotide tiếp nhận điều kiện nghiêm ngặt PCR, thay vì sử dụng hai primer đặc biệt như những primer tiêu chuẩn. Thành phần còn lại của phản ứng giống như bình thường, nhưng nội dung của chu trình thì hoàn toàn khác. Sau bước một, sự kiện mở dây đôi DNA của chu kỳ đầu tiên, nhiệt độ của phản ứng được phép giảm xuống khoảng 40°C, bấy giờ primer điều hành có thể tác động ở đầu dây đơn DNA, tại nhiều nơi trong genome, để bắt đầu tổng hợp DNA. Sau hai vòng khuếch đại trong điều kiện không nghiêm ngặt lắm (relaxed), người ta sử dụng PCR bình thường và người ta có thể quan sát các sản phẩm PCR không chuyên biệt ở trên gel. Nếu AP-PCR được lập lại, thì sản phẩm của PCR sẽ được sản xuất giống hệt nhau.

Thay vì sử dụng primer có 20 nucleotide, William và ctv. (1990) đã sử dụng primer có 10 nucleotide. Vì nhiệt độ T_m thấp hơn rất nhiều ở thể 10 mer so với 20 mer, cho nên William và ctv. đã sử dụng mọi điều kiện giống nhau cho tất cả các chu kỳ. Thể đa hình DNA được xác định thông qua sản phẩm PCR trên gel, với đa hình về độ dài của các đoạn PCR. William gọi đó là **RAPD**, viết tắt từ chữ Random Amplified Polymorphic DNA.

Bảng 4.3 : So sánh mức độ hiệu quả của RFLP và RAPD trong phân tích di truyền

Đặc điểm	RFLP	RAPD
Công lao động	15-23	6
Thời gian thực hiện	7 ngày	14 giờ
Số lượng mẫu DNA	1000 ng	25 ng
Chất lượng mẫu DNA	50+kb	10+kb
Alen	codominant	dominant
Số alen / locus	nhiều	hai
Hoạt động của marker	+/- ngẫu nhiên, không có trong vùng high copy	+/- ngẫu nhiên
Mức độ tìm thấy những	có	không
Loci tương đồng	(inter-taxa parisons)	(simplified polyploid genetics)

Tính chất có thể tái tạo của RAPD marker: tính chất này là điều quan tâm đặc biệt của các nhà chọn giống và di truyền. Có hai qui trình trắc nghiệm tính chất tái tạo này. Một là, thử hai hoặc ba mẫu hoàn toàn độc lập. Thí dụ như ly trích DNA ba lần từ một dòng, rồi chạy RAPD với ba mẻ DNA. Sau đó người ta tính được khả năng tái tạo của nó. Hai là, xét nghiệm qui luật phân ly Mendel trên một quần thể phân ly. Reiter và ctv. (1993) đã xác định có 225 RAPD marker không nằm trong 392 trường hợp phân ly theo Mendel.

Số bản sao của RAPD marker :

Nghiên cứu về chuỗi mã di truyền của RAPD marker, Reiter và ctv. (1993) đã xác định số bản sao (copy) của 18 RAPD marker có tính đa hình. Trong số đó, có 9 marker được lai với ít hơn 3 bản sao trong genome. Các RAPD còn lại được lai với các chuỗi mã di truyền có mặt tại 3-10 bản sao. Nếu RAPD được sử dụng như những chất thăm dò có tính lai (hybridization probes), lúc bấy giờ người ta sẽ xem nó như những RFLP marker, nó có thể phân lập thành thể đa hình có độ dài rất giới hạn.

Phương pháp này tùy thuộc vào điều kiện một đoạn 10mer vừa đủ cho một genome. Mục đích của nó là tăng hoạt nhiều lần segments của genome muốn có, theo kiểu ngẫu nhiên. Hai primer có thuật ngữ là "forward" và "reverse" ký hiệu F và R, sẽ đủ để tác động lẫn nhau, trên hai dây đối xứng nhau, nhằm tăng hoạt vùng bị can thiệp. Đối với DNA thực vật, các primer RAPD thường tăng hoạt 5-15 bands của 200-4000 bp xét theo chiều dài. Do đó mức độ "amplification" của nó rất tốt sau khi điện di, nhưng tần suất đa hình (polymorphism) khá hạn chế. Trong khi RFLP có tính chất "codominant" thì RFLP có tính chất "dominant", do đó những gen điều khiển tính trạng nào đó có tính lặn (recessive) sẽ khó tìm thấy sự đa hình trong điện di. Hơn nữa, RAPD có tính chất ngẫu nhiên, nên việc lập lại phân tích điện di để tìm liên kết gen và marker, thường không thống nhất (hình 4-8)

4.3.3. SSCP marker

SSCP được viết tắt từ chữ "single-strand conformation polymorphism". Chúng ta biết rằng ALP không phải luôn luôn được tìm thấy nếu những amplicon có cùng một độ dài, thậm chí trong trường hợp chúng có biến dị di truyền giữa những amplicon.

Người ta đã tìm thấy có sự chuyển dịch của đoạn DNA dạng dây đơn, ngắn, trong điều kiện chưa qua quá trình biến hóa DNA thành dây đơn (denaturation). Người ta giả định rằng: sự thay đổi chuỗi mã di truyền DNA là do sự thay đổi ngoại hình của dây đơn (single-strand conformation). Sự thay đổi này làm cho DNA chuyển dịch trên gel, tạo ra thể đa hình.

Trong phân tích SSCP, phản ứng chuẩn PCR đã hoàn thành. Sản phẩm của PCR này lại bị mở dây đơn lần nữa. Người ta ngâm các mẫu này trong nước đá. Bấy giờ hiện tượng snap-back sẽ xảy ra trên cấu trúc thứ cấp. Để tránh hiện tượng đứt gãy cấu trúc thứ cấp, các mẫu này phải được xử lý trong điều kiện lạnh. Nếu P³² được dùng trong PCR, thì phim

chụp X quang sẽ thể hiện rõ trên gel. Nếu không, người ta sẽ dùng bạc để nhuộm gel. Nhuộm bạc trên DNA dây đơn (SS DNA) sẽ nhạy cảm gấp trăm lần hơn nhuộm ethidium bromide.

4.3.4. STS marker

Sự chọn lọc giống nhờ marker [MAS = marker-assisted selection] có thể làm tăng hiệu quả lai tạo giống cây trồng đối với những tính trạng rất khó phân lập kiểu hình. Mặc dù phương pháp này đã xây dựng được bản đồ gần như khá đủ với DNA marker, nhưng việc ứng dụng của những nhà chọn giống vẫn phải bị lệ thuộc vào chi phí rất đắt của công nghệ gen.

Kỹ thuật DNA marker trên cơ sở vị trí được đánh dấu có tính chất mã di truyền : "STS" viết tắt từ chữ sequence-tagged sites, là một cách để khắc phục trở ngại nói trên.

Khái niệm STS do Olson và ctv (1989) đề xuất. Trong khi đánh giá hiệu quả PCR trên lĩnh vực nghiên cứu genome con người, họ ghi nhận những chuỗi mã DNA dạng copy đơn của một vị trí đã biết rồi trên bản đồ, có thể được xem như một marker, để lập bản đồ di truyền và bản đồ vật lý các gen quan trọng trong các nhiễm thể.

Thay vì bảo quản, duy trì vật liệu sinh học, các nhà khoa học bảo quản các thông tin có tính chất điện tử [electronic information], vì nó dễ nhận ra, sắp xếp theo thứ tự, và lan truyền. Những marker được lưu trữ có tính chất điện tử này đã được gọi là STS. Ứng dụng STS để phát triển các bản đồ vật lý từ những bản đồ di truyền, đang được tiến hành trong nhiều chương trình nghiên cứu, kể cả cây lúa (Inoue và ctv. 1994).

STS-based PCR và kỹ thuật marker phân tử :

Một STS là một đoạn ngắn của chuỗi mã di truyền, được tìm thấy bởi PCR (Saiki và ctv, 1985). Mỗi STS được ghi nhận trên bản đồ, tại một vị trí chuyên biệt nào đó như là một ranh giới (landmark) trong genome. Do vậy người ta còn dùng thuật ngữ standard landmarks để chỉ STS marker. Kỹ thuật PCR có vẻ thích hợp hơn kỹ thuật DNA blotting trong chọn lọc giống nhờ marker (MAS). Nó yêu cầu DNA có số lượng và chất lượng thấp hơn DNA blotting. Do đó qui trình ly trích DNA có thể được cải tiến đơn giản hơn, mà vẫn tránh được hiện tượng mẫu lớn. Nó còn có tính tự động hóa rất cao, không phải sử dụng phóng xạ, hay yêu cầu các hệ thống tìm kiếm có tính chất hóa sinh phức tạp (Zheng và ctv. 1995). STS-based PCR có ưu điểm là một bộ phận giản đơn, có tính sinh sản nhanh trên gel : agarose hoặc polyacrylamide. Bộ phận này dễ nhận thấy và dễ diễn giải, trong hầu hết các trường hợp nó là codominant. Những marker như vậy cho phép thể dị hợp có thể được phân biệt với hai thể đồng hợp.

STS-based PCR có hai nhược điểm: (1) nó cần có dữ liệu mã di truyền thích hợp đối với từng locus, (2) nó không có tính chất polymorphic như những DNA marker khác, thí dụ

như microsatellites (Wu và Tanksley 1993). Tính chất quan trọng của vấn đề thứ hai - tần suất đa hình thấp - thay đổi ở mỗi locus và ở mỗi cặp lai cụ thể.

Việc sáng tạo ra STS là tiền đề cho phép chúng ta chuyển đổi bản đồ di truyền thành bản đồ vật lý. Inoue và ctv. (1994) đã tìm ra các bộ PCR primer, mỗi primer chừng 20 base, để dò tìm các sequence của RFLP landmarker, chứa đựng các clone của genome cây lúa. Các marker này phân bố khắp 12 nhiễm thể, và được thể hiện như copy dạng đơn trong phân tích Southern. Với những PCR protocol được cải tiến hiện nay, người ta đã tạo ra được 63 standard STS landmarker. 63 STS trên cơ sở RFLP landmarker này phân bố trong genome cây lúa 4.3×10^8 bp, với khoảng cách trung bình của mỗi STS là 6.5×10^6 bp. Tóm lại, việc tạo ra STS sẽ theo trình tự như sau :

- Đọc sequence của RFLP qua phân tích với sự trợ giúp của computer
- Thiết kế primer F và R (theo kinh nghiệm, hoặc nhờ phần mềm computer)
- Sáng tạo ra STS, xét nghiệm thông qua PCR
- Thử nghiệm tìm ra enzyme thích hợp để có đa hình sau điện di, và có marker có tính chất codominant

Người ta có thể tạo ra STS bằng cách chạy sequence các clone, sau đó mapping chúng. Qui định này có tính bắt buộc đối với microsatellites, tiếp theo đó là những marker khác được sử dụng để chuyển đổi ra STS, như cDNA, gen được tạo thành clone, và các sản phẩm PCR được tạo thành clone. Số lượng lớn của ETS (expressed tagged site) của cDNA cây lúa cũng có thể tạo ra STS. Chuỗi mã có tên gọi là single-pass automated sequence cung cấp thông tin (350-450 bp) chỉ từ một đoạn cuối của mỗi clone. Điều này đáp ứng đầy đủ thông tin cho ETS được sử dụng để xây dựng bản đồ vật lý, hoặc để nghiên cứu tính chất tương tự trong nguồn dữ liệu, nhưng PCR amplicon này lấy từ ETS, thường rất ngắn để phát hiện polymorphism của bố mẹ, theo qui trình ALP (amplicon length polymorphism) hay PCR-based RFLP (Ghareyazie và ctv. 1995). Tuy nhiên người ta có thể khám phá polymorphism giữa những amplicon ngắn như vậy thông qua qui trình điện di đặc biệt (Fukuoka và ctv. 1994).

PCR-based marker :

Biến thiên di truyền giữa hai giống lúa có thể được phân tích nhờ khác biệt về độ dài của đoạn DNA đặc biệt nào đó thông qua kỹ thuật PCR, với cùng một cặp primer. Vì sản phẩm của PCR được gọi là amplicon, cho nên biến thiên như vậy được gọi là amplicon length polymorphism (ALP). Khi không có ALP nào được phát hiện bởi một cặp primer giữa hai giống, người ta áp dụng kỹ thuật PCR-based RFLP (viết tắt là PBR), với một restriction enzyme nào đó, cho thêm vào để tiêu hóa amplicon này. Việc xác định enzyme này cần làm với hàng loạt xét nghiệm, cho đến khi nào tìm được polymorphism xảy ra trên điện di.

Trước đó cặp primer (F và R) được thiết kế sau khi chạy sequence RFLP marker cần thiết. Việc thiết kế này tùy thuộc vào kinh nghiệm và kỹ năng của người làm việc trong phòng thí nghiệm.

Cả hai ALP và PBR đều được xem như là PCR-based marker

Thí dụ trong bảng 4-4 cho thấy các STS primer được dùng để tìm xem gen kháng có chứa trong bố mẹ và con lai. Trong trường hợp *Xa-21* và *Pi-12 (t)*, chúng ta không cần tiêu hóa với một enzyme tương ứng, nó vẫn cho đa hình, người ta gọi là ALP. Trường hợp *Pi-2* và *xa-5*, người ta phải sử dụng enzyme tương ứng để tiêu hóa sản phẩm PCR, nó mới cho kết quả đa hình mong muốn, người ta gọi đó là PBR (hình 4-10)

Bảng 4.4: STS được liên kết với các gen kháng bệnh đạo ôn và bạc lá của cây lúa

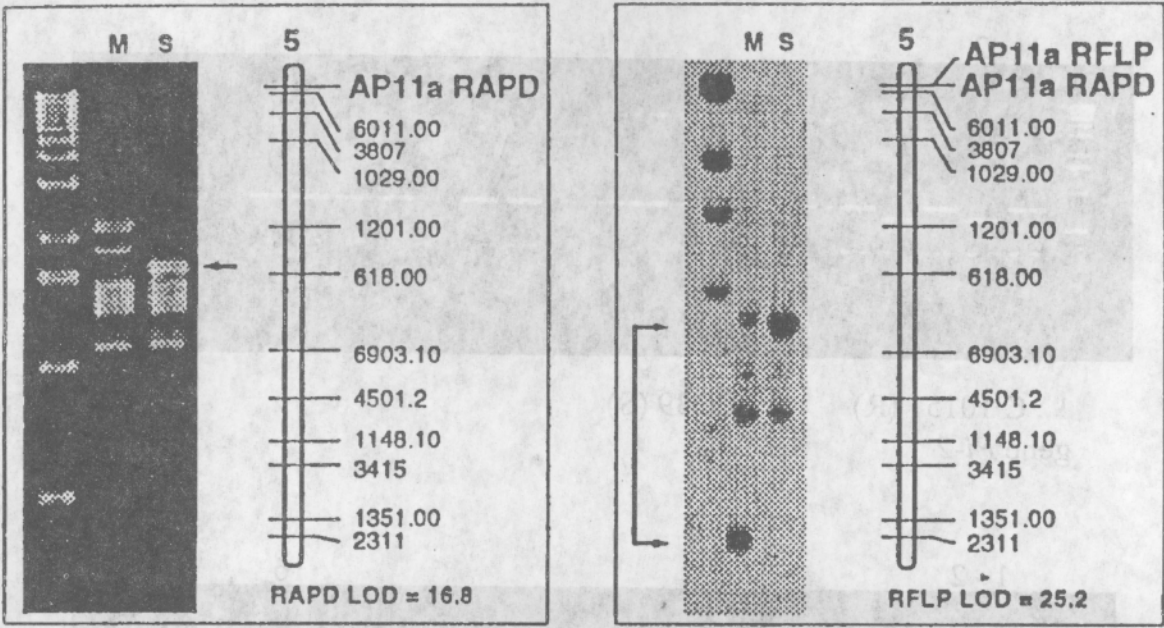
Marker	Gen liên kết	Nh. Thể số	STS primer		ALP / enzyme trong PBR
			5'	3'	
RG64	<i>Pi-2</i>	6	GTTGTTTGAGCTCTCCAATGCCTGTTC	CTGCAGTGCAATGTACGGCAAGG	<i>HaeIII</i>
pTA8	<i>Xa-21</i>	11	AGACGCGGAAGGGTGGTTC	CCCGGAAGCGCGGTGTAATCGAAAGATGAA	ALP
RG556	<i>xa-5</i>	5	TAGCTGCTGCCGTGTGC	AATATTTTCAGTGTGCATCTC	<i>DraI</i>
P265.560	<i>Pi-12(t)</i>	12	CAGCTGTTTCAGTCGTTTG	CAGCTGTTTCATACAAGAAAT	ALP

(Zheng và ctv. 1995)

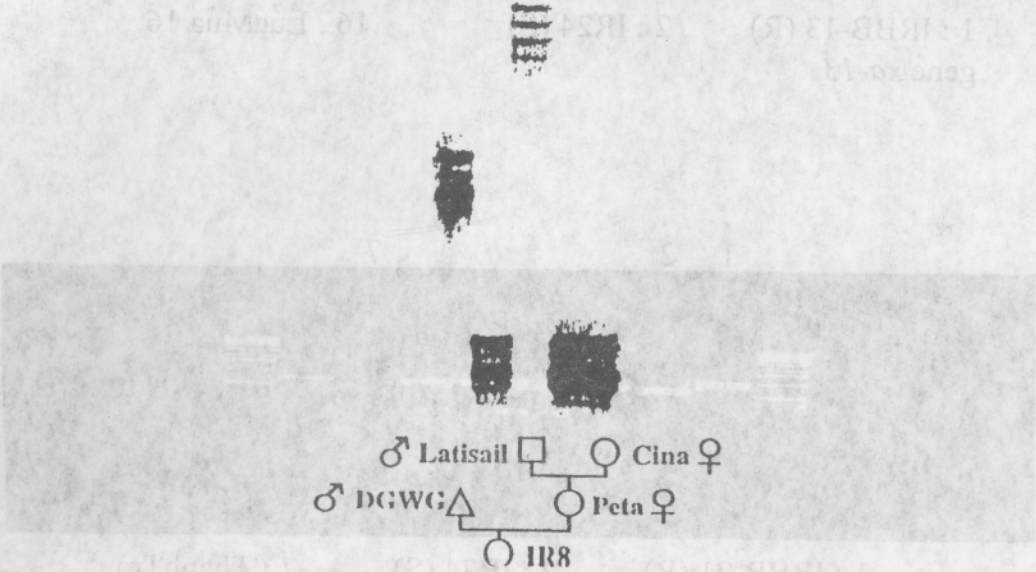
4.3.5. Microsatellite marker

Những marker vi vệ tinh (microsatellite) đã được ứng dụng khá thành công từ 1990 đến nay. Nó được dùng để phát hiện các bệnh ung thư thường gặp trên người. Nó cũng được dùng để phát hiện các gen có ích trong thực vật, sự liên hệ về huyết thống (hình 4-9)

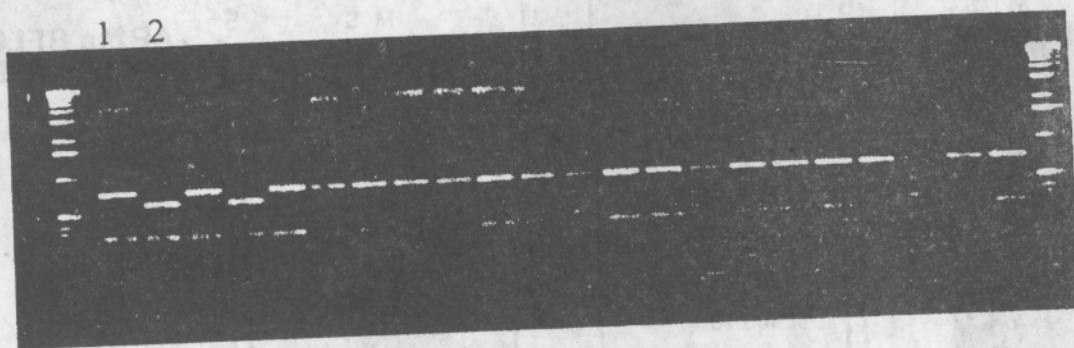
Những nghiên cứu đầu tiên về SSLP (simple sequence length polymorphism) cho thấy nó rất phong phú và phân bố rộng khắp ở genome cây lúa (Wu và Tanksley 1990, Panaud và ctv. 1995). Người ta thấy rằng: cần phải cải tiến để có những marker cung cấp mức độ cao hơn tính đa dạng alen, dùng trong thẩm định quỹ gen. Những marker vi vệ tinh đã đáp ứng được phần nào yêu cầu này. Bởi vì nó cho kết quả rất nhanh, tin cậy hơn RAPD, có thể nhuộm mẫu bằng bạc (Ag), không dùng chất đồng vị phóng xạ.



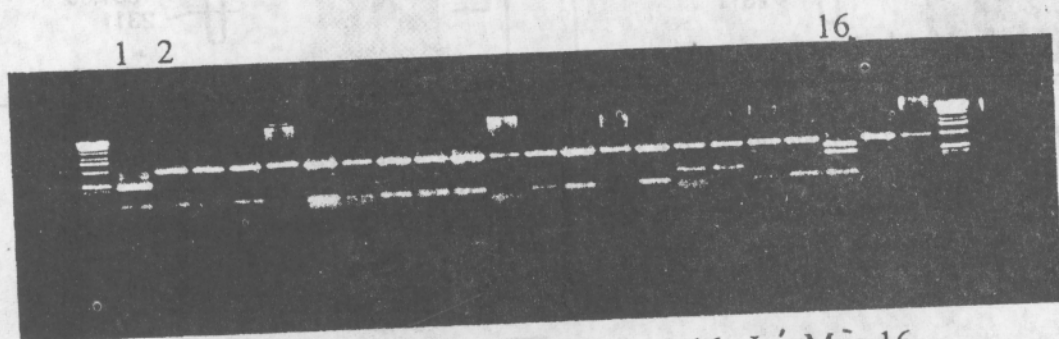
Hình 4.8 : Vị trí trên bản đồ của RAPD (AP11a) và tính tương xứng với RFLP của nó. Sự phân ly của AP11a [mũi tên], được cho thang điểm, và được xem xét theo vị trí của bản đồ RFLP trên đậu nành. Đặc điểm của RAPD marker là tạo ra nhiều băng, khuếch đại tốt (Tingey và ctv. 1992)



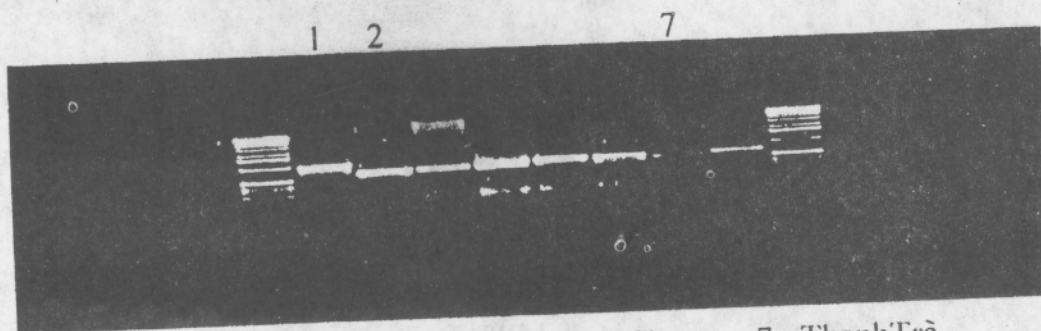
Hình 4.9 : Áp dụng microsatellite marker để nghiên cứu mối quan hệ huyết thống trên cây lúa. Sự đa dạng của marker RM21 trong gia phả giống lúa IR8 (Panaud và ctv. 1996)



1 : C 10151 (R) 2 : CO39 (S)
gene *Pi-2*



1 : IRBB-13 (R) 2 : IR24 (S) 16 : Lúa Mùa 16
gene *xa-13*



1 : IRBB-21 (R) 2 : IR24 (S) 7 : Thanh Trà
gene *Xa-21*

Hình 4.10 : Áp dụng STS marker để điều tra tính kháng bệnh đạo ôn (gen *Pi-2*), bệnh bạc lá gen (*xa-13*), và gen (*Xa-21*) của các giống lúa địa phương Việt Nam. R: giống kháng, S: giống nhiễm (Bửu và ctv. 1997)

Bảng 4.5 : Danh sách một vài microsatellite marker trên cây lúa

MARKER	PRIMER	CHUỖI MÃ GIẢN ĐƠN LẬP LẠI	Sản phẩm PCR kích thước (bp)
RM100	CATGGAGAGGAACTGGTGTT CTCTGATTTCTACCTCTCTC	GAA(GA) ₂₇ CATC	204
RM122	GAGTCGATGTAATGTCATCAGTGC GAAGGAGGTATCGCTTTGTTGGAC	TAA(GA) ₇ A(GA) ₂ A(GA) ₁₁ TTGC	227
RM123	GCCGTCTAACGTTTGTGGGTTA CACAGAAATGTGGAGCTCGAAC	AAC(GA) ₂₀ AGTAA	224
RM148	ATACAACATTAGGGATGAGGCTGG TCCTTAAAGGTGGTGCAATGCGAG	CTCTAT(GT) ₁₂ TTT	129
RM163	ATCCATGTGCGCCTTTATGAGGA CGCTACCTCCTTCACTTACTAGT	GGA(GGAGA) ₄ (GA) ₁₁ C(GA) ₂₀ GG	124
RM164	TCTTGCCCGTCACTGCAGATATCC GCAGCCCTAATGCTACAATTCTTC	(GT) ₁₆ TT(GT) ₄ GAG	246
RM167	GATCCAGCGTGAGGAACACGT AGTCCGACCACAAGGTGCCTTGTC	GGAA(GA) ₁₆ GGGG	128
RM168	TGCTGCTTGCCCTGCTTCCTTT GAAACGAATCAATCCACGGC	CGG(T) ₁₅ (GT) ₁₄ TTC	116
RM200	CGCTAGGGAATTTGGATTGA CGATGAGCAGGTATCGATGAGAA	CT(GA) ₁₆ GG	122
RM120	ACACAAGCCCTGTCTCACGACC CGCTGCGTCATGAGTATGTA	GT(GA) ₉ TAG(ATC) ₄	173
RM203	CCTATCCCATTAGCCAAACATTGC GACGCCAACCTGGAGTTAATTACC	TATT(AT) ₂₁ CCCCC	203
RM226	CGTCACGTTTACGGATTCCAGATC GTAGTTACCGGAATTGGTTCATCC	TAC(AT) ₄₁ GTCC	226
RM268	GTGCTATGCACGATCCATAGCA CGTTTCTTTGGAAGCGGAGGGA	(CCAN) ₇ N ₇₇ A ₁₃ GA ₃ G ₈ GA ₃ GA ₅	268

Marker vi vệ tinh bao gồm một chuỗi mã gốc (core sequence) được lập lại nhiều lần, và phân tán rộng khắp trong genome, trên nhiều loci, mỗi locus chứa alen thích ứng với mỗi dạng khác nhau về lần lập lại của nó (core repeat), và nó rất nhạy cảm (Ramakrishna và ctv. 1995). Do đó nó còn được gọi với thuật ngữ chuỗi mã giản đơn lập lại nhiều lần (simple sequence repeats = SSR). Motif có 15 bp repeat trong gen điều khiển protein III của thực khuẩn thể M13 đã được áp dụng rất có hiệu quả để khám phá đa hình và điều tra marker trong DNA

fingerprinting (Vassart và ctv. 1987, Ryskov và ctv. 1988). Marker vi vệ tinh chứa đựng các đoạn mã di truyền giản đơn, lặp lại nhiều lần là một công cụ rất có giá trị để phân tích di truyền, trong đánh dấu gen kháng rầy nâu ở cây lúa (Bửu và ctv. 1997).

Bảng 4.6 : Danh sách các dạng SSR (chuỗi mã giản đơn lặp lại) trong thực vật (Mohan và ctv. 1997)

Loài	SSR
Arabidopsis	AT, AG, GA, CT, TC, CA
Barley	AT, TCT, CT, TG, CTT, TGC, ATTT
Cam quýt	ATT, (TAT)(TAG)
Conifer	CA
Nho	GA, GT, A
Bắp	T, AG, CA, AC, AG, GA
Lúa	GA, GT, AT, GGT
Cao lương	AG, AC, AAC, AAG
Đậu nành	AT, ATT, TAT, CT, TA, AAT, TAA
Hướng dương	CA
Cà chua	GT, GA, ATT, GATA
Cây nhiệt đới	AC, A
Lúa mì	AA, CAG, GA, GT, CT
Khoai sọ	CT, AT, (TA)(CA)

4.3.6. AFLP marker

AFLP là chữ viết tắt của Amplified Restriction Fragment Length Polymorphism. Nó được áp dụng cho kỹ thuật DNA fingerprinting để tìm kiếm tính chất đa hình của DNA giữa các mẫu xét nghiệm khác nhau. Những sản phẩm fingerprint này có thể được xem như một công cụ nhằm xác định sự phân lập của một mẫu DNA đặc biệt nào đó. Những fingerprint cũng được dùng làm nguồn cung cấp các marker di truyền, hình thành bản đồ liên kết gen.

Kỹ thuật DNA fingerprinting đã và đang được phát triển khá thành công với hướng chiến lược nghiên cứu như sau :

- Fingerprinting trên cơ sở lai, có tính chất kinh điển : bao gồm kỹ thuật cắt DNA bằng restriction endonuclease tương ứng, kỹ thuật điện di, RFLP được phát hiện bởi kỹ thuật lai với probe thông qua xét nghiệm Southern.
- Fingerprinting trên cơ sở PCR: bao gồm kỹ thuật khuếch đại DNA *in vitro* bản

primer đặc biệt hay primer ngẫu nhiên, kỹ thuật polymerase trong máy thermocycler. Sản phẩm PCR được phân biệt nhờ điện di trên gel, được phát hiện nhờ phương pháp nhuộm hoặc phương pháp đánh dấu primer. Kỹ thuật đã được sử dụng theo chiến lược này là RAPD, DAF, AP-PCR (bảng 4-1).

Kỹ thuật AFLP là kỹ thuật DNA fingerprinting, kết hợp cả hai chiến lược này. Nó dựa trên cơ sở khuếch đại có chọn lọc (selective amplification) một subset các đoạn DNA đã bị cắt, sử dụng trong PCR. Mẫu DNA được tiêu hóa với restriction endonuclease, và adapters của dây đôi được gắn vào đầu dây DNA để phát sinh ra dây đơn template (mang mã ngược với dây gốc), sau đó cho khuếch đại nhờ kỹ thuật PCR. Chuỗi mã của những adapter và vị trí tương ứng của nó trên dây DNA hoạt động như một primer ở vùng gắn vào nhau, để liên tục khuếch đại các đoạn DNA. Những nucleotide có tính chọn lọc phát triển thành những restriction fragment, được bổ sung thêm vào đầu 3' của PCR primer (hình 4-11). Do đó chỉ có những subset của các đoạn DNA này mới được ghi nhận. Chỉ có những restriction fragments mà nucleotide của nó định vị xung quanh restriction site che khuất các nucleotide có tính chọn lọc, mới có thể được khuếch đại. Subset của những đoạn DNA được khuếch đại sẽ đem phân tích thông qua điện di trên polyacrylamide gel để có sản phẩm fingerprint.

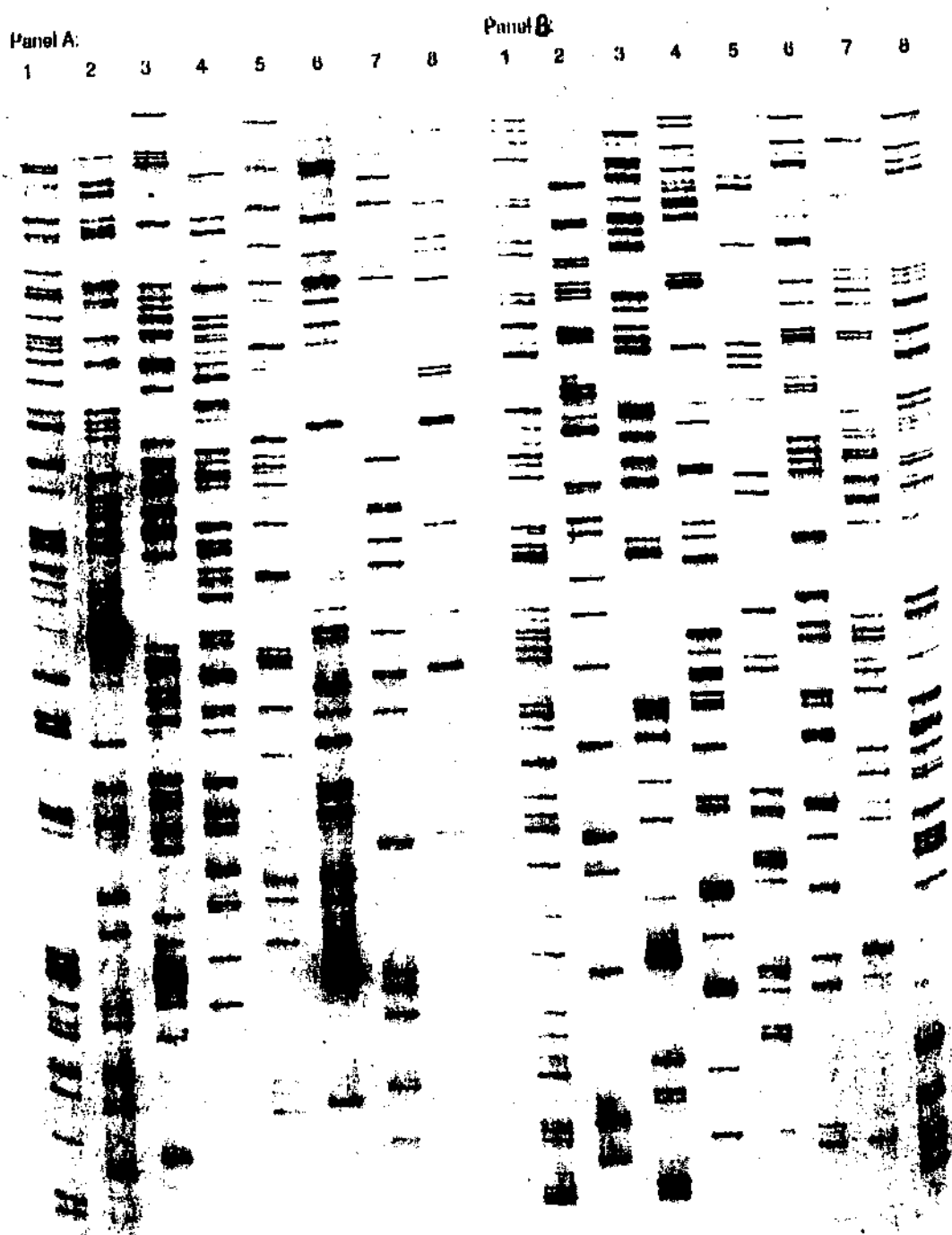
Người ta so sánh các DNA fingerprint này bằng cách quan sát các băng có tính đa hình hay không đa hình. Tính chất đa hình được phát hiện trên DNA fingerprint như vậy có thể cho biết sự thay đổi của chuỗi mã DNA bao gồm đột biến, hoặc tạo ra một restriction site mới, hoặc lấp đoạn, đảo đoạn, thiếu đoạn giữa hai restriction site. Nghiên cứu tính đa hình của DNA nhờ AFLP là mô hình mẫu trong nghiên cứu di truyền Mendel, do đó, nó có thể được dùng cho việc phân lập marker phân tử, xây dựng bản đồ gen (hình 4-11).

Sử dụng hai restriction endonuclease : *EcoRI* và *MseI*.

EcoRI có vị trí xác định là 6 bp, và *MseI*. Có vị trí xác định là 4 bp. Khi sử dụng chúng chung với nhau, những enzyme này sẽ tạo ra những đoạn DNA rất ngắn, chúng sẽ được khuếch đại khá tốt, và đạt được kích thước tối hảo (<1 kb) chạy điện di trên polyacrylamide gel. Do việc thiết kế primer và kỹ thuật khuếch đại, người ta thường dùng phối hợp *EcoRI*-*MseI* hơn là phối hợp *EcoRI*-*EcoRI* hoặc *MseI*-*MseI*.

Thành công của kỹ thuật AFLP tùy thuộc vào phản ứng của enzyme tiêu hóa có trọn vẹn hay không? Do đó, chúng ta phải thận trọng trong việc phân lập DNA có chất lượng cao, không làm tạp nuclease hoặc ức chế nó.

Số lượng C và G trong các nucleotide có tính chọn lọc sẽ rất quan trọng quyết định số fragments được khuếch đại. Thông thường, càng nhiều C và G có trong primer, thì các DNA fragments được khuếch đại càng ít. Tương tự như vậy, nếu kích thước genome càng nhỏ, số fragments được khuếch đại sẽ càng ít, và fingerprint càng đơn giản hơn rất nhiều.



Hình 4.11 : Áp dụng AFLP trong kỹ thuật fingerprinting để nghiên cứu DNA của *Arabidopsis*. Trong panel A, người ta sử dụng *EcoRI* primer [E-TR] theo sau đó là *MseI* primer : M-CAA [cột 1], M-CAC [cột 2], M-CAT [cột 3], M-CTC [cột 6], M-CTG [cột 7], và M-CTT [cột 8]. Trong panel B, người ta sử dụng *MseI* primer [M-CAG], theo sau là *EcoRI* primer : E-AA [cột 1], E-AC [cột 2], E-AT [cột 4], E-TA [cột 5], E-TC [cột 6], E-TG [cột 7], E-TT [cột 8]

Bảng 4.7 : Chuỗi mã của adapter và primer được dùng trong phân tích AFLP .

Adapter	Oligonucleotide sequence
<i>Pst</i> I adapter sequence 1 (96)	5' CTCGTAGACTGCGTACACATGCA-3'
<i>Pst</i> I adapter sequence 2 (97)	3' CATCTGACGCATGT-5'
<i>Mse</i> I adapter sequence 1 (A18)	5' GACGATGAGTCCTGAG-3'
<i>Mse</i> I adapter sequence 1 (A19)	3' TACTCAGGACTGAT-5'
Primers / preamplification:	
<i>Pst</i> I primer (96)	5' CTCGTAGACTGCGTACATGCA-3'
<i>Mse</i> I primer (H18)	5' GACGATGAGTCCTGAG-3'
Primers / khuếch đại chọn lọc:	
<i>Pst</i> I primer (90)	5' GACTGCGTACATGCACCA-3'
<i>Pst</i> I primer (91)	5' GACGTCGTACATGCAGTT-3'
<i>Pst</i> I primer (98)	5' GACGTCGTACATGCAGAC-3'
<i>Pst</i> I primer (99)	5' GACGTCGTACATGCATGG-3'
<i>Mse</i> I primer (F10)	5' GATGAGTCCTGAGTAACAC-3'
<i>Mse</i> I primer (F38)	5' GATGAGTCCTGAGTAAACC-3'
<i>Mse</i> I primer (F39)	5' GATGAGTCCTGAGTAACCA-3'
<i>Mse</i> I primer (F41)	5' GATGAGTCCTGAGTAACAA-3'
<i>Mse</i> I primer (G08)	5' GATGAGTCCTGAGTAAACG-3'
<i>Mse</i> I primer (G23)	5' GATGAGTCCTGAGTAACAG-3'
<i>Mse</i> I primer (G24)	5' GATGAGTCCTGAGTAACAT-3'
<i>Mse</i> I primer (G25)	5' GATGAGTCCTGAGTAACGA-3'
<i>Mse</i> I primer (G26)	5' GATGAGTCCTGAGTAACGT-3'
<i>Mse</i> I primer (G27)	5' GATGAGTCCTGAGTAACCT-3'

Protocol đầu tiên cho kỹ thuật AFLP do Zabeau và Vos (1993) nhấn mạnh đến sự quan trọng của việc làm thuần khiết hai enzyme sử dụng, sau đó là chuẩn bị adapter và thiết kế primer một cách thận trọng. Ngày nay nó đã được cải tiến khá nhiều với những công dụng như sau :

- Công cụ có hiệu quả phát hiện tính chất đa hình
- Xây dựng bản đồ di truyền có mật độ cao của genome
- Xây dựng bản đồ DNA marker có hiệu quả nhất so với những marker khác (Vos và ctv. 1995)
- Khám phá ra những clone của genome, thí dụ YAC
- Fingerprinting các đoạn DNA đã được cloned như cosmid, P1, BAC, YAC (Vos và ctv. 1995).

TÀI LIỆU THAM KHẢO

1. Bùi chí Bửu, K Renganayaki, AS Reddy. 1997. *Phân tích di truyền tính kháng rầy nâu của giống lúa hoang nhờ marker phân tử*. Tr. 79-82. Kết quả nghiên cứu khoa học 1977-1997. Viện Lúa Đồng Bằng Sông Cửu Long. Nhà XB Nông Nghiệp.
2. Buu BC, NV Tao, NT Lang. 1997. *Rice germplasm evaluation assisted by DNA markers in Mekong Delta - Vietnam*. OMon Rice 5: 96-98
3. Fukuoka S, T Inoue, A Miyao, L Monna, HS Zhong, T Sasaki, Y Min e. 1994. *Mapping of sequence-tagged sites in rice by single strand conformation polymorphism*. DNA Res. 1: 271-277
4. Ghareyazie B, N Huang, G Second, J Bennett. 1995. *Classification of rice germplasm. I. Analysis using ALP and PCR-based RFLP*. Theor. Appl. Genet. 91: 218-227
5. Inoue T, HS Zhong, A Miyao, I Ashikawa, L Monna, S Fukuoka, N Miyadera, Y Nagamura, N Kurata, T Sasaki, Y Min e. 1994. *Sequence-tagged sites (STSs) as standard landmarks in rice genome*. Theor. Appl. Genet. 89: 728-734
6. Michelmore RW, I Paran, RV Kesseli. 1991. *Identification of markers linked to disease resistance genes by bulked segregant analysis : a rapid method to detect markers in specific genomic regions by using segregant populations*. Proc. Natl. Acad. Sci. USA 88:9828-9832
7. Mohan M, S Nair, A Bhagwat, TG Krishna, M yano. 1997. *Genome mapping, molecular markers and marker-assisted selection in crop plants*. Molecular Breed. 3:87-103
8. Olson M, L Hood, C Cantor, D Botstein. 1989. *A common language for physical mapping of the human genome*. Science 245: 1434-1435
9. Panaud O, X Chen, SR McCouch. 1995. *Frequency of microsatellite sequences in rice (Oryza sativa L.)*. Genome 38:1170-1176
10. Panaud O, X Chen, SR McCouch. 1996. *Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (Oryza sativa L.)*. Mol. Gen. Genet. 252:597-607
11. Ramachrisna W, KV Chowdari, MD Lagu, VS Gupta, PK Ranjekar. 1995. *DNA fingerprinting to detect genetic variation in rice using hypervariable DNA sequences*. Theor. Appl. Genet. 90: 1000-1006

12. Reiter RS, JGK William, KA Fedman, Rafalski, SV Tingey, PA Scolnik. 1992. *Global and local genomic mapping in Arabidopsis thaliana by using recombinant inbred lines and random amplified polymorphic DNAs*. Proc. natl. Acad. Sci. USA 89:1477-1481
13. Ryskov AP, AG Jincharadze, MI Prosnayak, PL Ivanov, SA Limborska. 1993. *M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms*. FEBS Lett 233: 388-392
14. Saiki RK, DH Gelfand, S Stoffel, SJ Scharf, R Higuchi, GT Horn, KB Mullis, HA Ehrlich. 1988. *Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase*. Science 239: 487-391
15. Tanksley RD, S Ahn, T Fulton, SR McCouch, Z Yu, Z Wong, K Wu. 1990. *RFLP mapping of the rice genome*. 2nd Int. Rice Gen. Symp. P.58. IRRI, Philippines.
16. Tingey SV, JA Rafalski, JGK William. 1992. *Genetic analysis with RAPD markers*. Appl. Od RAPD Technol. To Plant Breed. 3-8
17. Vassart G, M Georges, R Monsieur, H Brocas, AS Lequarre, D Christophe. 1987. *A sequence of M13 phage000 detects hypervariable minisatellite0s in human and animal DNA*. Science 235:683-684.
18. Vos P, R Hogers, M Bleeker, M Reins, T van de Lee, M Hornes, A Frijters, J Pot, J Peteman, M Kuiper, M Zabeau. 1995. *AFLP: a new technique for DNA fingerprinting*. Nucleic Acids Res. 23 (21): 4407-4414
19. Welsh J, M McClelland. 1990. *Fingerprinting genomes using PCR with arbitrary primers*. Nucl. Acids Res. 8: 7213-7218
20. William JGK, AR Kubelic, KJ Livak, JA Rafalski, SV Tingey. 1990. *DNA polymorphism amplified by arbitrary primers are useful as genetic markers*. Nucleic Acids Res. 18:6531-6535
21. Wu KS, SD Tanksley. 1993. *Abundance, polymorphism and genetic mapping of microsatellites in rice*. Mol. Gen. Genet. 241: 225-235
22. Zabeau M, P Vos. 1993. *European Patent Application*, publication n°. EP 0534858
23. Zheng K, N Huang, J Bennett, GS Khush. 1995. *PCR-based marker-assisted selection in rice breeding*. IRRI Discussion Paper Series No. 12. IRRI, Philippines.

PHƯƠNG PHÁP TÍNH GIÁ TRỊ LIÊN KẾT GEN

5.1. KHOẢNG CÁCH GIỮA HAI GEN

Giả sử hai loci A và B có những alen A, a và B, b theo thứ tự, định vị trên cùng một nhánh của nhiễm thể. Recombination fraction c_{AB} giữa chúng là xác suất mà giao tử được truyền đi của một cá thể nào đó có tính chất recombinant (tái tổ hợp). Trong trường hợp không có ảnh hưởng xáo trộn nào khi phân ly, AB / ab (nhận giao tử AB và ab) được xem như chuyển đi bốn dạng giao tử như sau :

$$\frac{1-c_{AB}}{2} AB + \frac{c_{AB}}{2} Ab + \frac{c_{AB}}{2} aB + \frac{1-c_{AB}}{2} ab$$

Recombination fraction được xem như nằm trong quãng $[0, 1/2]$, đôi khi nó được chuyển thành thông số liên kết (linkage parameter), ký hiệu

$$AB = 1 - 2 c_{AB}$$

Gọi khoảng cách vật lý là d_{AB} giữa hai vùng bất kỳ nào của nhiễm thể, tính bằng kb, hoặc bp. Bản đồ gen trên cơ sở khoảng cách vật lý như vậy có thể được hình thành theo phương pháp có tên gọi là "radiation hybrid mapping", trong đó các dòng tế bào lai có chứa một đoạn nhiễm thể, được sản xuất bởi radiation breakage (đứt khúc do bức xạ), những dòng này được thanh lọc để tìm xem cái nào là hybrid định vị trên một locus đã biết trước.

Khoảng cách giữa hai gen được minh họa bằng thuật ngữ **genetic map distance** x_{AB} hoặc x . Đơn vị để tính khoảng cách này là centi Morgan (cM), đó là khoảng cách trong đó có một quần chéo (crossing over) được tìm thấy, trên một giao tử, trên một thể hệ.

Recombination fraction c có thể được tính theo khoảng cách x như sau :

$$c_{AB} = 1/2 (1 - e^{-2x})$$

Ghi chú: c_{AB} biến thiên từ 0 đến 0,5 trong khi x tiến tới 0. Viết lại phương trình:

$$x = -1/2 \ln(1 - 2c_{AB})$$

Kosambi (1944) phát triển thêm

$$x = \frac{1}{4} \ln \left(\frac{1 + 2c_{AB}}{1 - 2c_{AB}} \right)$$

5.2. DỰ ĐOÁN RECOMBINATION

5.2.1. Phương pháp hồi giao

Từ hai dòng đồng hợp tử AABB và aabb, con lai F1 hoàn toàn dị hợp AB/ab, và hồi giao AB/ab x aabb sẽ cho ra bốn nhóm con lai

Bảng 5.1 : Phân bố con lai thông qua hồi giao

Kiểu gen	Con lai				Total
	AB/ab	Ab/ab	aB/ab	ab/ab	
Kiểu hình	AB	Ab	aB	ab	n
Thực tế	a	b	c	d	n
Lý thuyết	n/4	n/4	n/4	n/4	n

Ở locus A, có a+b cá thể nhận alen A và c+d cá thể nhận alen a, do đó phép thử chi bình phương về lý thuyết phân ly độc lập của các alen sẽ là :

$$\chi_A^2 = \frac{\left(a + b - \frac{n}{2}\right)^2}{\frac{n}{2}} + \frac{\left(c + d - \frac{n}{2}\right)^2}{\frac{n}{2}} = \frac{(a + b - c - d)^2}{n}$$

$$\chi_B^2 = \frac{(a + c - b - d)^2}{n}$$

Đối với những cá thể nhận alen B, sự phân ly ở locus A được tính theo a - c, trong khi đó những cá thể nhận alen b được tính bằng b - d. Liên kết giữa những loci này làm cho những alen không được chuyển đi một cách độc lập, do đó sự phân ly ở một locus này bị ảnh hưởng bởi alen có mặt tại locus kia. Đối với locus A, ảnh hưởng có thể định lượng như sau :

$$(a - c) - (b - d)$$

Do đó liên kết gen được tìm thấy bằng cách so sánh giữa a + d với b + c

Giá trị chi bình phương :

$$\chi_{AB}^2 = \frac{(a + d - b - c)^2}{n}$$

Theo Bailey (1961), trong ba giá trị chi bình phương, mỗi giá trị có độ tự do là 1, như vậy tổng là 3 df (độ tự do) được áp dụng khi so sánh bốn nhóm con lai theo mẫu trên :

$$\chi^2 = \frac{(a - 1/4n)^2}{1/4n} + \frac{(b - 1/4n)^2}{1/4n} + \frac{(c - 1/4n)^2}{1/4n} + \frac{(d - 1/4n)^2}{1/4n}$$

$$= \chi_A^2 + \chi_B^2 + \chi_{AB}^2$$

Thí dụ: kết quả của Goodman và ctv (1980) trên enzymes *Mdh5* của bắp với alen *E12*, *E15*, và enzyme *Got3* với alen U4, U8

Test cross : $\frac{E12, U4}{E15, U8} \times \frac{E12, U4}{E12, U4}$

Con lai sẽ là : $\frac{E12, U4}{E12, U4} \quad \frac{E12, U8}{E12, U4} \quad \frac{E15, U4}{E12, U4} \quad \frac{E15, U8}{E12, U4}$

Giá trị thực tế ghi nhận được :

53	13	9	55	(total 130)
----	----	---	----	-------------

Giá trị lý thuyết sẽ là :

130/4	130/4	130/4	130/4
32,5	32,5	32,5	32,5

Phép thử chi bình phương :

$$\chi^2(Mdh5) = \frac{(66 - 64)^2}{130} = 0.03$$

$$\chi^2(Got3) = \frac{(62 - 68)^2}{130} = 0.28$$

$$\chi^2(Mdh5, Got3) = \frac{(108 - 22)^2}{130} = 56.89$$

$$\chi^2 = \frac{20.5^2 + 19.5^2 + 23.5^2 + 2^2 \cdot 52}{32.5} = 57.20$$

Cả hai loci phân ly một cách bình thường, nhưng có hiện tượng rất mạnh mẽ đối với sự liên kết giữa hai loci này.

Trong trường hợp những loci không liên kết nhau, có nghĩa là $c = 0,5$ thì mỗi dạng giao tử AB, Ab, aB, ab của một dị hợp tử AB/ab được dự đoán sẽ có tần suất bằng nhau.

Trong trường hợp những loci liên kết hoàn toàn, có nghĩa là $c = 0$, dự đoán bấy giờ là không có recombinant.

Khi số lượng recombination gia tăng, hai gamete AB và ab của bố mẹ, được dự đoán sẽ giảm tần suất, và Ab cũng như aB sẽ gia tăng tần suất.

Bảng 5.2 : Phân bố con lai thông qua hồi giao trong trường hợp loci liên kết

Kiểu gen	Con lai				Total
	AB/ab	Ab/ab	aB/ab	ab/ab	
Kiểu hình	AB	Ab	aB	ab	n
Thực tế	a	b	c	d	n
Lý thuyết	$n/2 (1-cAB)$	$n/2 cAB$	$n/2 cAB$	$n/2 (1-cAB)$	n

Đối với backcross AB / ab x ab / ab, giá trị lý thuyết phải bao gồm recombinant fraction cAB (theo bảng)

Giá trị Likelihood của cAB được tính :

$$L(cAB) \propto (1 - cAB)^{a+d}(cAB)^{b+c}$$

Chỉ số của likelihood thông qua hai yếu tố :

$$S_{cAB} = \frac{a + d}{1 - cAB} + \frac{b + c}{cAB}$$

$$l_{cAB} = \frac{n}{cAB(1 - cAB)}$$

$$cAB = \frac{b + c}{n} \text{ (giá trị MLE)}$$

Phương sai của MLE

$$\text{Var}(cAB) = \frac{cAB(1 - cAB)}{n}$$

Giá trị maximum likelihood phải bao hàm thông số này, sao cho c biến thiên trong quãng [0, 0,5]. Nếu tỉ số (b + c) / n lớn hơn 0,5 thì MLE được diễn giải như sau :

$$c = \min\left(\frac{b + c}{n}, 0,5\right)$$

5.2.2. Phương pháp phân tích quần thể F2

Bảng 5.3 : Quần thể F2

Kiểu gen	Con lai				Total
	AB/ab	Ab/ab	aB/b	ab/ab	
Kiểu hình	AB	Ab	aB	ab	n
Thực tế	a	b	c	d	n
Lý thuyết	$9n/16$	$3n/16$	$3n/16$	$n/16$	n

$$\chi_A^2 = \frac{(a+b-3/4n)^2}{3/4n} + \frac{(c+d-1/4n)^2}{1/4n}$$

$$= \frac{(a+b-3c-3d)^2}{3n}$$

$$\chi_B^2 = \frac{(a+c-3b-3d)^2}{3n}$$

$$\chi_{AB}^2 = \frac{(a+9d-3b-3c)^2}{9n}$$

Bảng 5.4 : Phân bố con lai F2 trong trường hợp loci liên kết với nhau

Kiểu gen	Con lai				Total
	AB/ab	Ab/ab	aB/b	ab/ab	
Kiểu hình	AB	Ab	aB	ab	n
Thực tế	a	b	c	d	n
Lý thuyết	$n/4(2+\theta)$	$n/4(1-\theta)$	$n/4(1-\theta)$	$n/4\theta$	n

Phân ly ở mỗi locus bây giờ là 3 : 1, thay vì 1:1 như trường hợp hồi giao

Để ước đoán recombination fraction của F2, Mather (1951) sử dụng $= (1 - c_{AB})^2$ để đưa vào công thức likelihood :

$$L(\theta) \propto (2 + \theta)^a (1 - \theta)^{b+c} (\theta)^d$$

$$S(\theta) = - \frac{n\theta^2 - (a - 2b - 2c - d)\theta - 2d}{(2 + \theta)\theta(1 - \theta)}$$

$$I(\theta) = \frac{n(1 + 2\theta)}{2\theta(1 - \theta)(2 + \theta)}$$

Giá trị MLE theo c_{AB} :

$$c_{AB} = 1 - (\theta)^{1/2}$$

$$\text{Var}(c_{AB}) = \left(\frac{d_{c_{AB}}}{d\theta} \right)^2 \text{Var}(\theta)$$

$$= \frac{(1 - \theta)(2 + \theta)}{2n(1 + 2\theta)}$$

Phương pháp Moments estimator của Ficher và Balmakund (1928), Immer (1930) sử dụng $K = ad/bc$:

$$\frac{\epsilon(a)\epsilon(d)}{\epsilon(b)\epsilon(c)} = \frac{\theta(2+\theta)}{(1-\theta)^2}$$

cho thấy θ có thể tìm

$$(K-1)\theta^2 - 2(K+1)\theta + K = 0$$

Immer (1930) cho θ một giá trị $K = ad/bc$

Vì estimator là phương trình có tính chất đồng hợp tử, nên người ta có thể tính phương sai :

$$\begin{aligned} \text{Var}(K) &= (\partial T / \partial a)^2 (2+\theta) + (\partial T / \partial b)^2 (1-\theta) + (\partial T / \partial c)^2 (1-\theta) + (\partial T / \partial d)^2 \theta - (\partial T / \partial n)^2 \\ &= \frac{2\theta(2+\theta)(1+2\theta)}{n(1-\theta)^5} \end{aligned}$$

Tuy nhiên,

$$\text{Var}(K) = (dK/d\theta)^2 \text{Var}(\theta)$$

và

$$dK/d\theta = 2(1+2\theta)/(1-\theta)^3$$

Do đó :

$$\text{Var}(\theta) = 2\theta(1-\theta)(2+\theta)/n(1+2\theta)$$

Bây giờ, kết quả được ghi nhận giống như MLE của θ . Khám phá này không có lợi ích gì, trong đó, MLE được biết là những estimator có hiệu quả nhất, có nghĩa là nó có phương sai nhỏ nhất. Rất hiếm có một estimator nào khác với mức độ hiệu quả như vậy.

Gọi cAB_m và cAB_f là những recombination fraction của bố và mẹ, theo thứ tự :

$$\theta = (1 - cAB_m)(1 - cAB_f)$$

5.2.3. LOD scores

Haldane và Smith (1947) đề xuất phương pháp có thuật ngữ là "maximize the likelihood" $L(c)$ để tính recombination fraction. Để chứng minh hai loci có liên kết với nhau hay không, người ta so sánh giá trị likelihood của c value nào đó, có likelihood của $c = 0,5$ (loci không liên kết). Logarithm thập phân của tỉ số này được gọi là LOD score (log - odds) $Z(c)$:

$$Z(c) = \log \left[\frac{L(c)}{L(0,5)} \right]$$

5.3. THỰC HÀNH

- Để ước đoán recombination fraction giữa những nhóm esterase loci trong cây barley (lúa mạch), người ta dùng các cặp lai giữa những dòng cận giao. Con lai F1 được lai để sản xuất ra thế hệ F2 (intercross generation). Những tổ hợp lai này chứa đựng các alen thuộc nhiều loci khác nhau. Đối với loci A và B, người ta dùng 3 tổ hợp lai, và con lai F2 được quan sát ở family số 1, số 2 và số 6, theo thứ tự. Dùng tất cả số liệu để ước đoán recombination fraction giữa loci A và B. Chúng ta hãy tìm likelihood là kết quả của tất cả những likelihood của 9 family.

Bố mẹ	F1	Con lai	Family					
			1	2	3	4	5	6
$A_2A_2B_1B_1$	$A_2A_3B_1B_2$	$A_2A_2B_1B_1$	12					
$A_3A_3B_2B_2$		$A_3A_3B_2B_2$	15					
		$A_2A_3B_1B_2$	35					
$A_2A_2B_1B_1$	$A_2A_3B_1B_3$	$A_2A_2B_1B_1$	16	46				
$A_3A_3B_3B_3$		$A_3A_3B_3B_3$	33	32				
		$A_2A_3B_1B_3$	44	105				
		$A_2A_2B_1B_3$	3	0				
		$A_2A_3B_1B_1$	2	0				
		$A_2A_3B_3B_3$	2	0				
$A_2A_2B_1B_1$	$A_2A_3B_1B_2$	$A_2A_2B_1B_1$	71	90	92	81	80	95
$A_3A_3B_2B_2$		$A_3A_3B_2B_2$	69	99	108	60	57	80
		$A_2A_3B_1B_2$	138	158	164	142	118	137
		$A_2A_3B_1B_1$	1	0	0	0	0	0
		$A_3A_3B_1B_2$	1	0	0	0	0	0
		$A_2A_3B_2B_2$	1	0	0	0	0	0

Theo dự đoán, khi mức độ recombination khá nhỏ, thì các dị hợp tử được hình thành hoàn toàn từ giao tử của bố mẹ. Có 3 nhóm F2 : 1136 đồng hợp tử được hình thành từ sự phối hợp của giao tử bố mẹ, 1041 dị hợp tử được hình thành từ sự phối hợp của giao tử bố mẹ và 10 được hình thành từ một bố (mẹ) với một giao tử có tính chất recombinant.

Likelihood đối với recombination fraction c sẽ là :

$$L(c) = \alpha \left[\left(\frac{1-c}{2} \right)^2 \right]^{1136} 2 \left[\left(\frac{1-c}{2} \right)^2 \right]^{1041} 2 \left[\left(\frac{c(1-c)}{2} \right) \right]^{10}$$

$$\ln L(c) = \text{constant} + 4364 \ln(1-c) + 10 \ln(c)$$

$$\frac{\partial L(c)}{\partial c} = -\frac{4364}{1-c} + \frac{10}{c}$$

$$\text{Như vậy: } c = 10/4374 = 0,0023$$

Ghi chú :

$$\text{dị hợp tử} = 35+44+105+138+158+164+142+118+137 = 1041$$

đồng hợp tử =

$$12+15+16+46+33+32+71+90+92+81+80+95+69+99+108+60+57+80 = 1136$$

$$\text{một bố(mẹ) và một giao tử recombinant} = 3+2+2+1+1+1 = 10$$

$$4374 = (1136 \times 2) + (1041 \times 2) + (10 \times 2)$$

$$\text{Như vậy } A_{AB} = 1 - 2c_{AB} \text{ (linkage parameter)}$$

$$= 1 - (2 \times 0,0023) = 0,9954$$

2. Hai dòng bố mẹ cận giao có kiểu gen AABB và aabb với A trội so với a, và B trội so với b. Quần thể F2 có x, y, z, w là số cá thể có kiểu hình của AB, Ab, aB, ab. Trắc nghiệm liên kết gen giữa loci A và B khi x=30, y=15, z=14, w=1. Điều gì sẽ xảy ra nếu x=30, y=17, z=13, w=0 ?

Con lai F1 sẽ là dị hợp tử AB / ab

$$\frac{1-c}{2} AB + \frac{c}{2} Ab + \frac{c}{2} aB + \frac{1-c}{2} ab$$

Gọi x, y, z, và w là kiểu hình của AB, Ab, aB, ab

$$x + y + z + w = n$$

$$L(c_{AB}) \propto \left[\frac{3 - c_{AB} + c_{AB}^2}{4} \right]^x \left[\frac{c_{AB}(2 - c_{AB})}{4} \right]^{y+z} \left[\frac{(1 - c_{AB})^2}{4} \right]^w$$

Nên nhớ rằng, $c_{AB} = 0$ khi $y = z = 0$

Kiểu hình và kiểu gen ở F2

Kiểu hình	Kiểu gen	Tần suất
AB	AABB	$[(1 - cAB)/2]^2$
	AABb	$2[(1 - cAB)/2] [cAB/2]$
	AaBB	$2[(1 - cAB)/2] [cAB/2]$
	AaBb	$2[(1 - cAB)/2]^2 + 2[cAB/2]^2$
	AAbb	$[(cAB/2)]^2$
	Aabb	$2[(1 - cAB)/2] [cAB/2]$
	aaBB	$[(cAB/2)]^2$
	aaBb	$2(1 - cAB)/2] [cAB/2]$
ab	aabb	$[(1 - cAB)/2]^2$

Derivative thứ nhất của likelihood là :

$$\frac{\partial \ln L(cAB)}{\partial cAB} = \frac{-2x(1 - cAB)}{3 - 2cAB + c^2 AB} + \frac{2(y + z)(1 - cAB)}{cAB(2 - cAB)} + \frac{-2w}{1 - cAB}$$

Giá trị ước đoán maximum likelihood của cAB trong quãng (0, 0.5) là L(cAB)

Derivative thứ nhất sẽ trở thành zero nếu cAB biến thiên rất chặt trong quãng giữa 0 và 0,5

Ta có $\theta = c^2 AB$. Cho nên, likelihood sẽ là :

$$L(\theta) \propto (2 + \theta)^x (1 - \theta)^{y+z} \theta^w$$

Derivative thứ nhất của log-likelihood là zero khi:

$$n\theta^2 - (x - 2y - 2z - w) \theta - 2w = 0$$

$$cAB = (\theta)^{1/2}$$

Khi $x = 30, y=15, z=14, w=1$, giá trị quadratic của là

$$60\theta^2 + 29\theta - 2 = 0$$

Giải phương trình

$$\theta = 0,06121 \quad \text{và} \quad cAB = 0,2474$$

Trắc nghiệm chi bình phương

$$\chi^2 = (x - 3y - 3z + 9w)^2 / 9n = 4,27$$

Giá trị này có ý nghĩa, cho thấy ít tái tổ hợp hơn $cAB = 0,5$. Ở đây, liên kết gen đã được chứng minh

Nếu $w = 0$, chúng ta sẽ không thấy thể lặn kép xuất hiện, giá trị derivative sẽ là

$$\frac{\partial \ln L(cAB)}{\partial cAB} = -2(1 - cAB) \left(\frac{x}{3 - 2cAB + c^2 AB} - \frac{n - x}{cAB(2 - cAB)} \right)$$

Nếu $x < 2n / 3$, giá trị này không thể bằng zero, derivative sẽ dương, và maximum likelihood của cAB là 0,5.

Nếu $x = 2n / 3$, thì derivative = 0, khi $cAB = 0,5$

Trắc nghiệm chi bình phương

$$\chi^2 = (x - 3y - 3z + 9w)^2 / 9n = 6,67$$

Kết quả có ý nghĩa, cho thấy nhiều tái tổ hợp hơn $cAB = 0,5$. Ở đây liên kết gen không được chứng minh.

3. Kiểu gen của locus gây bệnh DD, DN, và NN có xác suất f_{DD} , f_{DN} và f_{NN} đối với kiểu hình xuất hiện bệnh. Tìm xác suất của bố mẹ có M_1M_2 marker genotype chuyển qua M_1D hay M_1N . Xác suất để M_1M_2 parent chuyển M_1 vào con lai bị bệnh?

Trong trắc nghiệm này :

$$\Pr(T : M_1, NT : M_2 \setminus T : D) = m(1 - m) + (1 - c - m)D / p$$

Tương tự :

$$\Pr(T : M_1, NT : M_2 \setminus T : N) = m(1 - m) + (1 - c - m)D / (1 - p)$$

Nếu $\Pr(T : M_1, NT : M_2)$ là xác suất để M_1M_2 father chuyển M_1 vào con lai bị bệnh (child affected)

$$\begin{aligned} \Pr(T : M_1, NT : M_2) &= \frac{\Pr(T : M_1, NT : M_2, \text{child affected})}{\Pr(\text{child affected})} \\ &= \frac{\Pr(T : M_1, NT : M_2 \setminus T : D)[f_{DD}p^2 + f_{DN}p(1 - p)]}{f_{DD}p^2 + f_{DN}p(1 - p) + f_{NN}(1 - p)^2} \\ &\quad + \frac{\Pr(T : M_1, NT : M_2 \setminus T : N)[f_{DN}(1 - p) + f_{NN}p(1 - p)^2]}{f_{DD}p^2 + f_{DN}p(1 - p) + f_{NN}(1 - p)^2} \end{aligned}$$

Xác suất để M_1M_3 mother chuyển M_1 vào affected child phải có trong kiểu gen của con lai. Nếu con lai là DD chẳng hạn, cả bố và mẹ phải cùng chuyển M_1D

M1	M2	M3	M4	M5	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
1	1	1	1	1	2	2	1	1	2	2	2	1	1	1
2	2	2	1	1	2	2	1	1	1	2	2	2	1	2
2	2	2	1	2	1	2	1	1	2	2	2	1	1	1
2	1	1	2	1	2	2	1	1	1	2	2	1	2	1
2	1	1	2	1	2	1	1	1	2	2	1	2	1	1
2	1	1	2	1	1	1	2	2	2	1	1	2	2	2
2	2	2	2	2	1	2	2	1	2	2	2	2	2	1
2	2	2	2	2	1	1	2	2	2	2	1	1	2	1
2	1	1	2	1	2	2	1	1	1	2	2	1	1	1
2	2	2	1	1	1	1	1	1	1	1	1	2	2	2
1	1	2	2	2	1	1	1	2	2	2	2	1	2	1
1	1	1	2	1	2	2	1	1	1	1	1	1	2	2
1	1	2	2	2	1	2	1	1	2	1	2	2	1	2
1	1	2	2	2	1	2	2	1	2	2	2	2	1	1
2	2	1	2	1	1	1	1	2	2	2	2	1	1	1
1	1	2	2	2	2	1	1	2	1	1	1	2	2	2
2	1	1	2	1	1	2	1	2	2	2	1	1	2	2
2	2	1	1	1	1	1	2	2	2	1	1	2	2	1
1	2	2	2	2	1	1	2	2	2	2	2	1	2	1
2	1	1	2	1	2	2	1	1	1	1	2	2	1	2
2	1	1	1	1	2	2	1	1	1	1	1	1	2	2
2	1	1	2	1	2	2	1	1	1	2	2	1	1	1
2	2	1	1	1	1	1	1	1	1	2	2	1	1	2
1	1	2	2	2	1	2	2	1	2	1	1	1	2	1
1	2	1	1	2	1	1	2	2	2	2	2	1	2	1
2	2	1	1	1	1	1	1	1	1	2	1	2	1	2
2	2	2	1	2	2	2	2	2	1	1	2	2	1	2
1	2	2	1	2	1	1	2	2	2	2	2	2	1	1
1	1	2	2	2	1	1	2	2	2	2	2	2	1	2
					2	2	1	1	1					

Nếu con lai là DN, một parent chuyển M_1D và một parent chuyển M_1N

4. Có hai dòng hoàn toàn cận giao P1 và P2 được xem xét đối với những alen luân phiên nhau tại 5 marker loci. Con lai F1 được hồi giao với P1 các genotype của 5-locus marker trong 100 member của quần thể hồi giao được ghi nhận theo bảng sau (1 có nghĩa là P1 genotype, 2 có nghĩa là F1 genotype). Vị trí theo thứ tự của những marker không được biết. Hãy thiết lập bản đồ di truyền của những marker này.

Recombination fraction được dự đoán cho từng cặp loci là tỉ số của time 12 hoặc 21

	1	2	3	4	5
1	-	.27	.70	.61	.81
2	.27	-	.57	.78	.60
3	.70	.57	-	.43	.23
4	.61	.78	.43	-	.42
5	.81	.60	.23	.42	-

Giá trị sar cho tất cả $5! / 2 = 60$ phần tử được trình bày theo bảng dưới đây. Giá trị cho thấy thứ hạng có sar nhỏ nhất là 12354.

Thứ tự					sar	Thứ tự					sar
1	2	3	4	5	1.69	2	3	1	4	5	2.30
1	2	3	5	4	1.49*	2	3	1	5	4	2.50
1	2	4	3	5	1.71	2	3	4	1	5	2.42
1	2	4	5	3	1.70	2	3	5	1	4	2.22
1	2	5	3	4	1.53	2	4	1	3	5	2.32
1	2	5	4	3	1.72	2	4	1	5	3	2.43
1	3	2	4	5	2.47	2	4	3	1	5	2.72
1	3	2	5	4	2.29	2	4	5	1	3	2.71
1	3	4	2	5	2.51	2	5	1	3	4	2.54
1	3	4	5	2	2.15	2	5	1	4	3	2.45
1	3	5	2	4	2.31	2	5	3	1	4	2.14
1	3	5	4	2	2.13	2	5	4	1	3	2.33
1	4	2	3	5	2.19	3	1	2	4	5	2.17
1	4	2	5	3	2.22	3	1	2	5	4	1.99

Thứ tự					sar					Thứ tự					sar				
1	4	3	2	5	2.21	3	1	4	2	5	2.69								
1	4	3	5	2	1.87	3	1	5	2	4	2.89								
1	4	5	2	3	2.20	3	2	1	4	5	1.87								
1	4	5	3	2	1.83	3	2	1	5	4	2.07								
1	5	2	3	4	2.41	3	2	4	1	5	2.77								
1	5	2	4	3	2.62	3	2	5	1	4	2.59								
1	5	3	2	4	2.39	3	4	1	2	5	1.91								
1	5	3	4	2	2.25	3	4	2	1	5	2.29								
1	5	4	2	3	2.58	3	5	1	2	4	2.09								
1	5	4	3	2	2.23	3	5	2	1	4	1.71								
2	1	3	4	5	1.82	4	1	2	3	5	1.68								
2	1	3	5	4	1.62	4	1	3	2	5	2.48								
2	1	4	3	5	1.54	4	2	1	3	5	1.98								
2	1	4	5	3	1.53	4	2	3	1	5	2.86								
2	1	5	3	4	1.74	4	3	1	2	5	2.00								
2	1	5	4	3	1.93	4	3	2	1	5	2.08								

Lập một dãy số chỉ cần có 5 số hạng theo thứ tự, nhưng mỗi dãy cần 4 decision. Thứ tự này được xếp như sau, và xác định đáp số là 12354

Bắt đầu	1st	2nd	3rd	4th	sar
1	12	412	4123	54123	1.87
2	12	123	1235	12354	1.49
3	23	235	2354	23541	1.83
4	45	453	1453	14532	2.20
5	35	354	2354	12354	1.49

Hiện nay, người ta đã thiết lập phần mềm trên computer để tính toán khoảng cách di truyền, vẽ bản đồ liên kết gen và marker. Chương trình có tên là MAP MARKER thường được cài đặt trong Macintosh.

TÀI LIỆU THAM KHẢO

5. Fisher RA, B Balmakund. 1928. *The estimation of linkage from the offspring of selfed heterozygotes*. J. Genet. 20:79-92
6. Haldane JBS, CBA Smith. 1947. *A new estimate of the linkage between the genes for colour-blindness and haemophilia in man*. Eugen 14: 10-31
7. Immer FR. 1930. *Gormulae and tables for calculating linkage intensities*. Genetics 15: 81-98
8. Kosambi DD. 1944. *The estimation of map distances from recombination values*. Ann. Eugen. 12:172-175
9. Mather K. 1951. *The measurement of linkage in heridity*. Second Edition. Methuen, London.
10. Weir BS. 1996. *Genetic Data Analysis II*. Sinauer Associates, Inc., USA.445p.

Chương VI

KỸ THUẬT NHÂN BẢN VÔ TÍNH (CLONING)

6.1. PHƯƠNG PHÁP CLONING

Bất cứ một qui trình cloning nào cũng phải có bốn phần căn bản như sau: (i) phương pháp tạo ra các đoạn DNA, (ii) phản ứng kết nối DNA và vectơ để tạo ra recombinant DNA, (iii) phương tiện du nhập recombinant nhân tạo này vào tế bào chủ, ở đó nó có thể tự tái bản được, (iv) phương pháp chọn lọc hoặc thanh lọc dòng vô tính nào đó của tế bào nhận, tế bào này tạo ra được thể recombinant này. Trong các phần trước, chúng ta đã đề cập đến qui trình phân cắt, kết nối nhờ enzyme, đặc tính của thể phage và vectơ plasmid với các yếu tố kiểm soát sự chọn lựa, giữa những phương pháp cắt và kết nối khác nhau, và những vectơ phân tử khác nhau. Sự lựa chọn này tùy thuộc vào loại hình nào của clone mà ta mong muốn, thí dụ như cDNA hoặc clone của genomic DNA.

6.2. KHO LƯU TRỮ DNA CỦA GENOME (LIBRARY)

Có hai nhóm kho lưu trữ (library): lưu trữ DNA của genome và lưu trữ cDNA (complementary DNA). Trong mục 6.2 này, chúng ta chỉ đề cập đến kho lưu trữ DNA của genome, vai trò của *E. coli* đối với việc xây dựng library trong vectơ, và kỹ thuật chromosome walking.

6.2.1. Nguyên tắc chung

Thí dụ chúng ta muốn clone một gen đơn nào đó của genome con người, chúng ta có thể cho phản ứng tiêu hóa toàn bộ DNA với một restriction endonuclease như *EcoRI* chẳng hạn. Sau đó chúng ta gắn các đoạn DNA này vào một vectơ thực khuẩn thể λ thích hợp. Tiếp theo đó, chúng ta cố gắng phân lập clone mà chúng ta mong muốn. Câu hỏi đặt ra là: có bao nhiêu recombinant mà chúng ta phải thanh lọc để phân lập ra cái đúng nhất? Giả sử *EcoRI* phân cắt ra một số đoạn DNA có độ dài trung bình 4 kb, và genome đơn bội của người có $2,8 \times 10^6$ kb, như vậy chúng ta có thể quan sát 7×10^5 các recombinant có tính độc lập (Old và Primose 1995). Những recombinant này phải được xử lý và thanh lọc để có một cơ hội tốt nhất, chứa đoạn mã di truyền mà ta mong muốn. Nói cách khác, chúng ta phải thu nhận một số lượng lớn các recombinant, là một bộ sưu tập đầy đủ của tất cả các DNA sequence trong genome con người. Một bộ sưu tập như vậy để từ đó chúng ta có thể rút ra một clone mà chúng ta mong muốn, được gọi là "gene library" (kho lưu trữ gen) hay ngân hàng gen.

Vấn đề nảy sinh trong kỹ thuật này là gen mong muốn có thể bị cắt mất một hoặc nhiều lần bởi *EcoRI*, cho đến nỗi nó không được nhận như là một đoạn phân tử đơn. Nhất là trong trường hợp gen có kích thước quá lớn. Nó có thể hi vọng được nhận biết qua các vùng kế cận của gen (flanking), hoặc vùng kế cận của toàn bộ nhóm gen (gene cluster). Fragment có độ dài khoảng 4 kb không được xem là ngắn. Vấn đề này có thể được khắc phục bằng cách cloning những DNA fragment có tính chất ngẫu nhiên, có kích thước khoảng 20 kb. Vì DNA là những đoạn ngẫu nhiên, cho nên chúng ta sẽ không có một sự loại bỏ ra có tính chất hệ thống của bất cứ một chuỗi mã nào. Hơn nữa, những clone này sẽ che lấp (overlap) lên những clone khác, tạo ra cơ hội để walking (thuật ngữ này có nghĩa là di chuyển theo kiểu dò dẫm) từ một clone này sang một clone khác kế cận. Kích thước của mỗi DNA được clone hóa càng lớn, thì clone được yêu cầu càng ít, trong một library hoàn chỉnh, hay gần như hoàn chỉnh. Câu hỏi đặt ra là: người ta cần phải có bao nhiêu clone? Gọi n là kích thước của genome có liên quan đến đoạn phân tử đơn sẽ clone hóa. Ở genome người có $2,8 \times 10^6$ kb, và kích thước của đoạn cần clone hóa là 20kb, thì $n = 1,4 \times 10^5$. Số recombinant có tính độc lập cần có trong kho lưu trữ phải lớn hơn n , bởi vì biến thiên của mẫu sẽ làm cho việc gắn vào các chuỗi mã rất nhiều lần, và việc loại ra chuỗi mã khác trong kho lưu trữ của n recombinant cũng nhiều lần. Clarke và Carbon (1976) đã đề xuất một công thức có liên quan đến xác suất P của bất cứ một chuỗi mã DNA nào trong một kho lưu trữ ngẫu nhiên của N recombinant có tính độc lập như sau :

$$N = \frac{\ln(1 - P)}{\ln(1 - 1/n)}$$

Do đó, để có xác suất 95% ($P=0,95$) ghi nhận sequence đặc biệt mong muốn trong kho lưu trữ DNA của người, với kích thước đoạn phân tử là 20 kb, thì số N sẽ là :

$$N = \frac{\ln(1 - 0,95)}{\ln(1 - 1/[1,4 \times 10^5])}$$

$$N = 4,2 \times 10^5$$

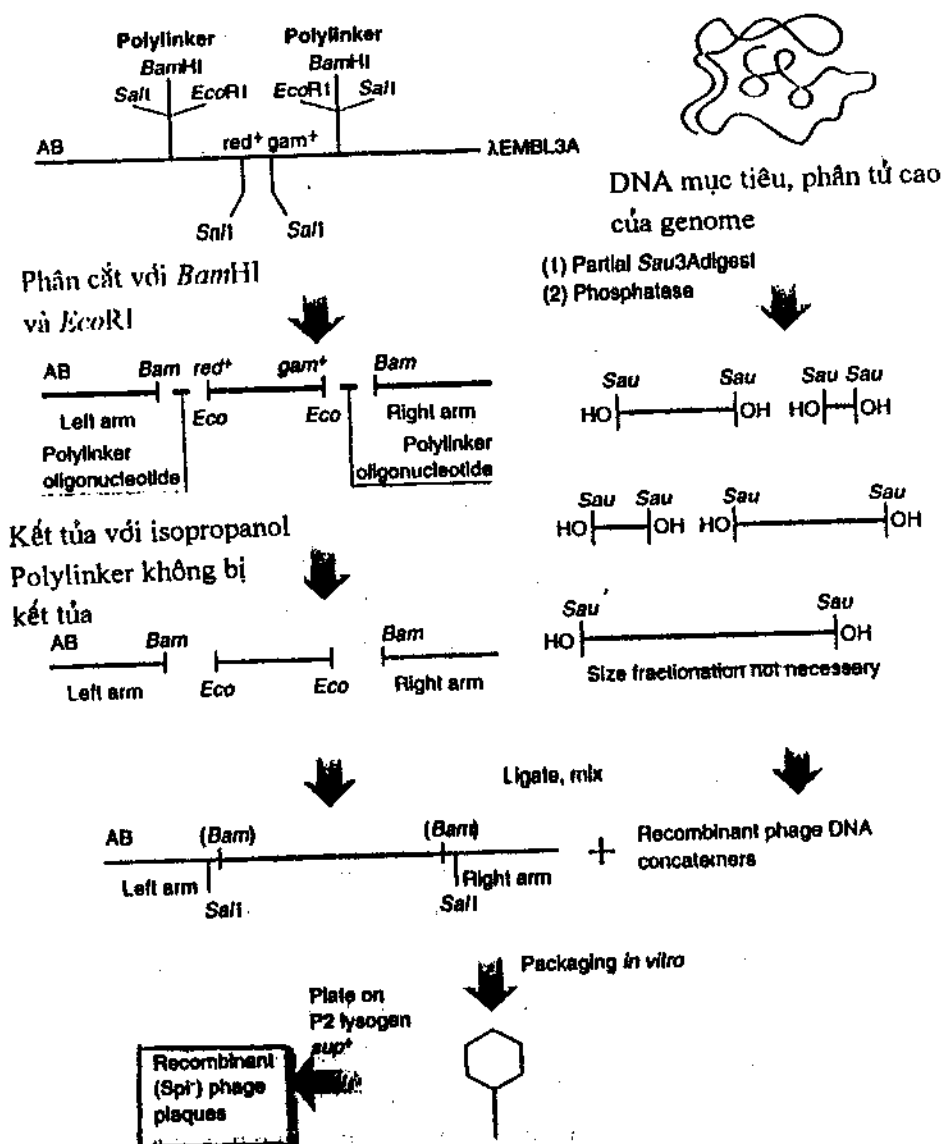
Ghi chú : Số recombinant cần có cao hơn ở mức xác suất 99% sẽ là $N = 6,5 \times 10^5$.

Người ta có nhiều phương pháp để sản xuất ra các đoạn phân tử có tính chất ngẫu nhiên, Thí dụ cho phân cắt ngẫu nhiên bằng cách tạo đứt gãy DNA bằng cơ học (mechanical shearing), nhưng cách thông dụng nhất là sử dụng restriction endonuclease (Maniatis và ctv. 1978). DNA cần nghiên cứu được tiêu hóa bởi sự trộn lại hai "enzyme phân cắt hạn chế". Những enzyme này có vị trí xác nhận của nó thuộc dạng tetranucleotide. Do đó, nó có mặt thường xuyên trong DNA này và trong sự tiêu hóa kép có tính giới hạn (limit double-digest). Điều này làm cho tiến trình sản xuất ra các sản phẩm nhỏ hơn 1 kb. Phản ứng restriction này chỉ được thực hiện để kéo dài có tính giai đoạn nào đó, sao cho việc tập hợp các fragments tương đối lớn, biến thiên khoảng 10-30 kb. Đây là một set rất có tính ngẫu nhiên và có hiệu quả

của những đoạn có tính chồng lắp lên nhau. Nó cũng có thể được phân biệt nhờ ly tâm trong một gradient của đường sucrose, hay điện di trên gel. Từ đó, chúng ta có thể tìm thấy một quần thể ngẫu nhiên của các đoạn phân tử có kích thước khoảng 20 kb, thích nghi cho việc gắn vào vectơ thực khuẩn thể lambda. Theo phương pháp của Maniatis, việc sử dụng hai restriction endonuclease khác nhau, có vị trí xác nhận thuộc dạng không tương hợp nhau, *HaeIII* và *AluI*, sẽ cho ra các đoạn gần như ngẫu nhiên (không hoàn toàn). Những enzyme này đều cho ra các blunt end (xem chương II), và phương pháp tạo clone như vậy sẽ cần thêm các linker (linker là phân tử DNA nhân tạo được gắn vào DNA mục tiêu, để có một chuỗi mã xác định chuyên biệt cho một enzyme phân cắt hạn chế nào đó). Một cách làm đơn giản hơn là sử dụng một enzyme như *Sau3AI*. Nó sẽ tạo ra một sự tiêu hóa từng phần, ít có tính chất ngẫu nhiên hơn phương pháp dùng một cặp enzyme nói trên. Tuy nhiên, có một thuận lợi lớn là các đoạn phân tử của *Sau3AI* rất sẵn sàng gắn vào vectơ thực khuẩn thể lambda, thí dụ như λ EMBL3, mà trước đó nó đã được tiêu hóa bởi *BamHI* (hình 6-1). Bởi vì *Sau3AI* và *BamHI* đều tạo ra đầu dây theo kiểu sticky (cohesive). Phương pháp tiêu hóa từng phần, như vậy, trong những recombinant của thực khuẩn thể lambda, đã được ứng dụng rộng rãi, nhằm hình thành kho lưu trữ DNA của genome.

Cosmid cũng có thể được chọn để thay thế vectơ λ thực khuẩn thể. Cosmid có hiệu quả cao và khả năng tốt hơn vectơ λ thực khuẩn thể (Old và Primrose 1995). Tuy nhiên, trong thực tế, người ta thấy rằng thanh lọc kho lưu trữ những recombinant của λ thực khuẩn thể bằng phương pháp plaque hybridization (lai đánh dấu) cho kết quả rõ ràng hơn thanh lọc library của vi khuẩn có chứa cosmid recombinant bằng phương pháp lai khuẩn lạc (colony hybridization). Đối với thực khuẩn thể, những quần thể DNA recombinant đầu tiên đóng kín trong host của nó. Người ta có thể thanh lọc ngay trong giai đoạn này. Những plates có chứa các recombinant được đánh dấu có thể được rửa và cho một kho lưu trữ có tính khuếch đại của recombinant phage. Kho lưu trữ này được bảo quản, nhờ phage có chu kỳ sống dài. Đối với khuẩn lạc của vi khuẩn có chứa cosmid cũng có thể được tồn trữ trong kho lưu trữ dạng khuếch đại (amplified) (Hanahan và Meselson 1980), nhưng quần thể vi khuẩn không thể được tồn trữ như quần thể phage.

Trong khi sử dụng các kho lưu trữ có tính khuếch đại, chúng ta cần chú ý một hiện tượng được gọi là "distortion" (sự mở xoắn). Không phải tất cả các recombinant trong quần thể đều được phát triển tốt như nhau, nghĩa là có sự biến thiên về kích thước DNA, về chuỗi mã. Điều này gây ảnh hưởng mạnh mẽ đến khả năng tự tái bản của recombinant phage, plasmid hay cosmid. Khi một kho lưu trữ nào đó đang trong tình trạng giai đoạn khuếch đại, thì những recombinant có thể sẽ gia tăng, hoặc giảm tần suất của nó, hoặc không tăng không giảm. Việc tạo ra các vectơ mới và xây dựng phương pháp cải tiến trong cloning đã làm cho công tác lập kho lưu trữ ngày càng đơn giản hơn, xác suất rủi ro ít hơn rất nhiều, như YAC, BAC, P1.



Hình 6.1 : Tạo ra một kho lưu trữ DNA của genome, sử dụng vectơ thực khuẩn thể λ [EMBL3A]. DNA phân tử cao được phân cắt bởi *Sau*3A. Những đoạn phân tử này được xử lý bằng phosphatase để loại gốc 5'-phosphate của nó ra ngoài. Vectơ được phân cắt bằng *Bam*HI và *Eco*RI trong khu vực của polylinker. Các đoạn phân tử cực nhỏ của polylinker giới hạn bởi *Bam*HI/*Eco*RI được loại ra nhờ phản ứng kết tủa với isopropanol. Những cánh tay của vectơ được kết nối với đoạn phân tử DNA đã được phân cắt. Dùng phosphatase để ngăn ngừa các DNA của genome nối trở lại với nhau, chỉ cho phép nó nối với vectơ. Vectơ không có DNA của genome gắn vào (non recombinant) không thể có vì các đoạn polylinker đều được kiểm soát, chỉ còn những phân tử recombinant phage. Có hai phage [thể đánh dấu] trên P2 lysogen của *sup*⁺ *E. coli*. Chọn lọc Sp⁺ sẽ đảm bảo sự hồi phục của phage thiếu những gen *red* và *gam*. Người ta rất cần một ký chủ của *sup*⁺. Bởi vì trong ví dụ này vectơ mang thể đột biến trong gen A và gen B. Chúng có thể được áp dụng để chọn lọc các recombinant (Old và Primrose 1995).

6.2.2. *E. coli* đối với xây dựng thư mục trong vectơ thực khuẩn thể

Người ta hi vọng : kho lưu trữ DNA của genome trong trường hợp phage λ (vectơ) sẽ chứa đựng hầu hết những chuỗi mã của genome. Tuy nhiên, người ta thấy rằng trong thực tế, genome của sinh vật eukaryotic chưa được tìm thấy các sequence của nó có trong kho lưu trữ này một cách phổ biến. Người ta còn quan sát hiện tượng mất đoạn (đứt đoạn) cũng có thể xảy ra trong khi cloning phân tử DNA của loài động vật có vú. Chỉ có một vài trường hợp đứt đoạn này có thể được bảo vệ tránh khỏi, nhờ phát triển trên sinh vật chủ (host) có tính chất thiếu tái tổ hợp (recombination-deficient) của *E. coli* *rec A*.

Leach và Stahl (1983) đã chứng minh rằng những sequence có tính lặp lại, đảo (có palindrome lớn, nghĩa là độ dài chuỗi 5'-3' lớn) sẽ không thể được clone hóa trong phage λ , trừ khi các dòng sinh vật chủ đặc biệt nào đó được sử dụng. Wyman và etv. (1985) đã phát triển thêm thí nghiệm này bằng cách sử dụng sinh vật chủ với thể đột biến là những gen *recB*, *recC* và *sbcB*, có chức năng tái tổ hợp. Những gen này phát triển nhiều recombinants trong phage λ có chứa những đoạn phân tử DNA của người, thì sẽ không phát triển được trên những dòng sinh vật chủ khác.

6.2.3. Vectơ là YAC

Trong kỹ thuật dùng YAC làm vectơ, có bốn loại khác nhau tùy theo tương tác giữa nó và tế bào nhận:

- YIp : viết tắt từ chữ yeast integrating plasmids
- YRp : viết tắt từ chữ yeast replicating plasmids
- YEp : viết tắt từ chữ yeast episomal plasmids
- YAC : viết tắt từ chữ yeast artificial chromosomes

Trong đó nhiễm thể nhân tạo bằng men được xem như có hiệu quả nhất. Những plasmids dây thẳng mang các cấu trúc có tính chức năng tại hai đầu của nó, được biết là telomer (Murray và Szostak 1983). Nó chứa đựng một kiểu liên hoàn chuỗi mã có tính chất lặp lại, hoạt động như một vệ tinh 5'-C₁₋₈(T/A)₁₋₄-3' có trong sinh vật thuộc eukaryote bậc thấp (Blackburn 1985). Plasmid dây thẳng có chứa gen *LEU2* là một marker có tính chọn lọc, cạnh bên là *CEN3* và *ARS1* trên đoạn cuối của *Tetrahymena* rDNA có kích thước 0,7 kb. Marker này không ổn định so với những plasmid vòng có kích thước ngắn hơn 10-16 kb. Nhiễm thể nhân tạo có thể không hữu hiệu trong việc thể hiện protein lạ trong men, tuy nhiên nó có thể cho phép một sự khai thác có tính chất hệ thống trên cơ sở : đặc điểm nhiễm thể, chức năng của men, loài trong eukaryote. YAC là vectơ được dùng trong việc thể hiện gen, trong kho lưu trữ cDNA nhiều hơn trong kho lưu trữ DNA của genome.

6.2.4. Vector là BAC

Trong chương III, chúng ta đã đề cập khá chi tiết về BAC. Ứng dụng BAC trong kỹ thuật cloning mới được phổ biến rộng rãi trong những năm gần đây, do những ưu điểm của nó.

- Cloning trong *Bacillus subtilis* mà prototype của nó là *Bacilli*.
- Cloning trong *E. coli*.
- Cloning trong *Streptomyces* và phage vector của nó đã được thảo luận bởi nhiều nhà nghiên cứu làm cơ sở cho việc ứng dụng BAC trong kỹ thuật cloning (Bailey và ctv. 1984, Bibb và ctv. 1985, Dhaese và ctv. 1984, Ehrlich và ctv. 1982, Gray và Chang 1981, William và ctv. 1981, Yansura và Henner 1984).

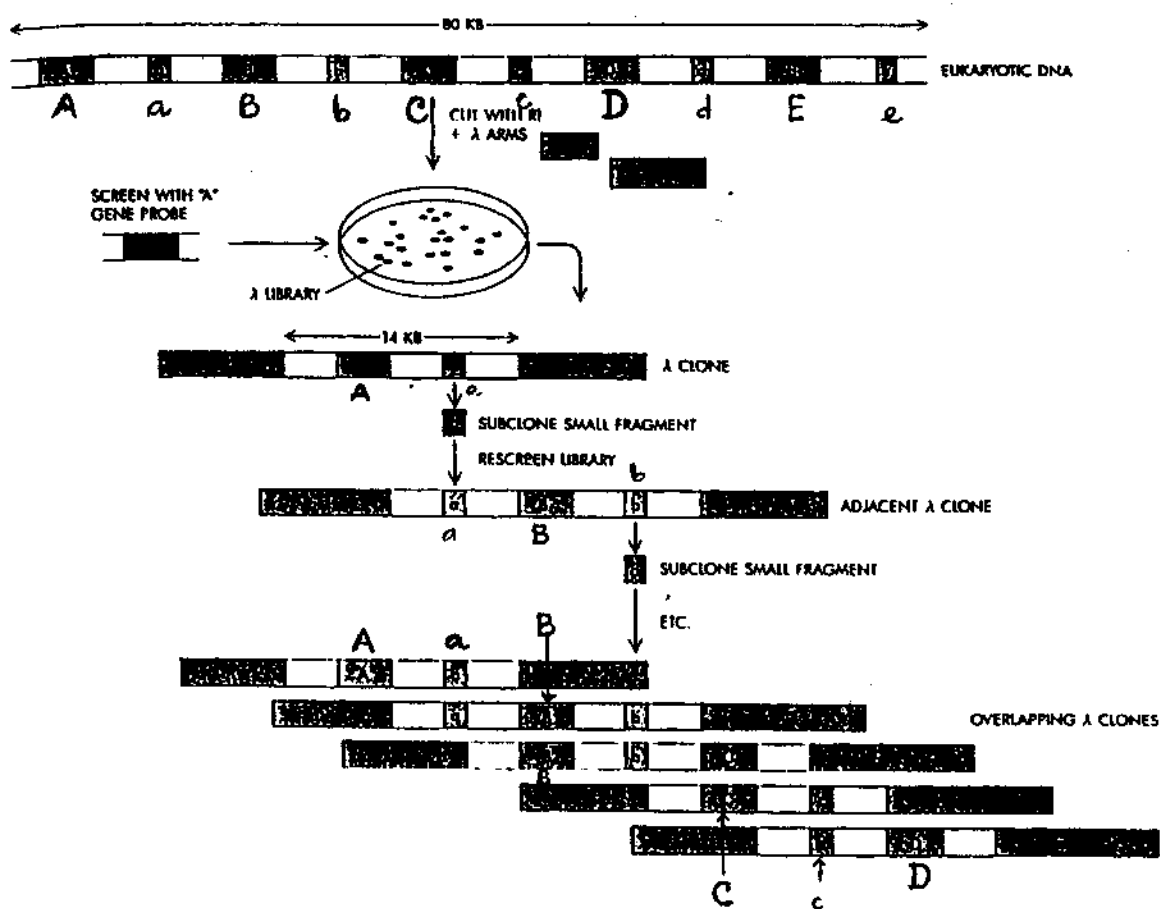
6.2.5. Chromosome walking

Thuật ngữ "walking" dọc theo nhiễm thể được dùng để mô tả một phương pháp cho phép chúng ta *phân lập các chuỗi mã của gen có chức năng chưa biết rõ, nhưng vị trí di truyền của nó thì rõ*. Thuật ngữ "chromosome walking" có tính chất quốc tế, nên chúng ta có thể sử dụng luôn không nhất thiết dịch sang tiếng Việt.

Nguyên tắc:

- Một đoạn phân tử được clone hóa phải có "điểm khởi đầu" được xác định để tiến trình walking bắt đầu. Trong genome của người, điểm bắt đầu có thể là một chuỗi mã protein của RFLP liên kết chặt với một locus điều khiển một bệnh nào đó. Trong ruồi *Drosophila*, điểm bắt đầu có thể được tìm thấy về tế bào học, nhờ lai *in-situ* đối với các nhiễm thể ở giai đoạn polytene.
- Một bộ các DNA được clone hóa, có tính chất ngẫu nhiên trong genome, được xác định vị trí, và một bộ khác được chọn, mà vị trí của nó trên nhiễm thể chưa được biết, nhưng người ta nghi rằng nó có thể ở gần nhất đối với vị trí trên bản đồ các thể đột biến đã nghiên cứu trước.
- Thư mục DNA của genome được thanh lọc lại với những clone được chọn (còn gọi là vật thể thăm dò = probe) để phân lập các clone khác có chứa DNA. Chúng phản ứng với nhau, tạo ra hiện tượng chồng lấp lên nhau (overlap). Hình overlap có thể ở bên trái hoặc ở bên phải.
- Hiện tượng tự tái bản của tiến trình walking này, xuất hiện dọc trên nhiễm thể.

Thí dụ : muốn tìm một gen B ở cách gen A theo dự đoán thông qua các xét nghiệm trước đây. Gen A đã được biết và được xem như điểm bắt đầu cho tiến trình walking. Gen A được clone hóa có tên là clone 1. Sequence của clone 1 có thể được dùng làm probe để phân lập clone 2, 3, 4. Clone 4 có thể được dùng làm probe để phân lập clone 5, và tiếp tục như vậy. Walking cho đến khi nào đụng gen B (hình 6-2). Những clone 1, 2, 3, 4, .. là những đoạn DNA ngẫu nhiên (dài 20 kb), có sẵn trong thư mục.



Hình 6.2 : Kỹ thuật chromosome walking. Một thực khuẩn thể lambda recombinant được ghi thực từ một kho lưu trữ của genome thuộc nhóm eukaryote, có thể được sử dụng để phân lập một recombinant khác kế cận nó trong genome. Recombinant đầu tiên được tạo dòng phụ (subclone) thành một mẫu nhỏ trong pBR322. Một đoạn phân tử DNA có tính chất subcloned là một chuỗi mã đơn, đoạn phân tử này khác với thể thăm dò (probe) dùng trong phân lập "recombinant" đầu tiên. Subclone này được dùng để thanh lọc lại kho lưu trữ λ , nhằm ghi nhận một recombinant nào đó, chồng lấp lên một phần của clone λ đầu tiên và DNA kế cận.

Walking có thể xảy ra theo cả hai hướng dọc trên nhiễm thể. Trong một bản đồ có qui mô khá lớn tạo ra nhờ walking, nó có thể được thành lập bằng cách hợp lại các clone có tính chất kế cận nhau.

Sự hiện hữu của những chuỗi mã có tính lặp lại là một khả năng đáng ghi nhận. Vì chúng có thể phân tán rải rác ở nhiều vị trí trong genome, làm đột phá tiến trình có tính chất thứ tự của hiện tượng walking. Cho nên chất thăm dò được sử dụng qua từng bước, từ một clone này sang một clone khác phải là một clone có sequence thật đồng nhất, hoặc là một subclone có một sequence đồng nhất.

Điều quan trọng là làm sao theo dõi ghi nhận sự kiện walking đang tiến gần tới mục

tiêu mong muốn. Trong chương trình áp dụng walking trong genome của người, người ta đã phân lập gene điều khiển sự hóa xơ nang, định vị trên nhiễm thể số 7. Công trình được tiêu chuẩn hóa như giáo trình trong việc tạo ra clone có chứa gen này (Rommens và ctv. 1989, Riordan và ctv. 1989, Kerem và ctv. 1989).

- Phát hiện những chuỗi mã sinh ra do lai chéo trong các loài khác. Điều này có quan hệ với việc bảo quản những sequence của gen trong quá trình tiến hóa, mà spacer hiện chưa được bảo quản tốt.
- Phân lập các tiểu đảo CpG (CpG islands), chuỗi mã có nhiều cặp CG
- Xuất hiện sự chuyển mã trong mô bình thường, mô này bị ảnh hưởng do bệnh trạng nói trên
- Phân lập cDNA tương ứng với đoạn phân tử trong mô bình thường này.
- Xác định việc mở khung đọc mã trong chuỗi mã nucleotide của đoạn phân tử này.

Về nguyên tắc, kỹ thuật chromosome walking rất đơn giản, nhưng khi "walking" trên một vùng rộng lớn, nó đòi hỏi nhiều kỹ thuật kèm theo. Walking trên khoảng cách lớn thường kết hợp với kỹ thuật có thuật ngữ là "chromosome jumping" (nhảy). Jumping sẽ khắc phục được tình trạng đoạn phân tử có tính lặp lại cao, không thể clone hóa, hoặc không ổn định (Rommens và ctv. 1989). Chromosome walking trở thành một kỹ thuật đóng vai trò then chốt trong dự án phân tích genome con người. Nó cũng rất quan trọng trong nghiên cứu genome ruồi *Drosophila*. Một trong những ứng dụng đầu tiên của kỹ thuật này là DNA được clone hóa từ loci *Ace* và *rocy* và nhóm gen *Bithorax* đồng dạng (homeotic) của ruồi *D. melanogaster* (Bender và ctv. 1983).

6.3. PHÂN TÍCH CHUỖI MÃ DNA CÓ PHÂN TỬ LỚN VÀ CHROMOSOME WALKING VỚI KHOẢNG CÁCH RỘNG

Kỹ thuật cloning cổ điển không thể cho phép chúng ta thực hiện walking trên hàng nghìn kb. Người ta ước lượng 1000 kb sẽ tương ứng với một khoảng cách cho tái tổ hợp là 1cM (1% tái tổ hợp) ở người. Các vấn đề trong phân tích bản đồ và walking, với khoảng cách lớn như vậy đã được khắc phục nhờ những tiến bộ như sau:

- Khám phá ra những restriction endonuclease, kết hợp phản ứng methylase và endo nuclease, cắt DNA ở vị trí rất không bình thường (infrequent).
- Phát triển điện di trên gel trong xung điện trường (pulsed-field) và trong đảo trường (field-inversion), đối với các DNA có phân tử quá lớn.
- Phát triển phương pháp chromosome jumping.

Người ta có xu thế sử dụng YAC làm phương tiện trong trường hợp cloning những đoạn phân tử DNA rất lớn như vậy.

6.3.1. Cắt DNA ở vị trí rất hiếm gặp

Những restriction endonuclease thuộc type II có chuỗi mã ghi nhận thuộc nhóm octanucleotide được thử nghiệm để cắt DNA ở vị trí không bình thường. Thí dụ như enzyme *NotI* (GCGGCCGC), *SfiI* (GGCCNNNNNGGCC), và *PacI* (TTAATTA). Những enzyme khác có chuỗi mã ghi nhận lớn tương tự, sẽ được khám phá thêm trong tương lai. Có những restriction endonuclease với chuỗi mã xác nhận ngắn hơn, nó thể hiện tính chất chuỗi mã ít thông dụng (uncommon), tìm thấy trong genome người. Thí dụ như *NruI* (TCGCGA) và *BssHII* (GCGGCG). Những chuỗi mã hoạt động này có hai CG dinucleotide. Chính CG dinucleotide rất hiếm thấy trong genome người (ngoại trừ tiểu đảo CpG), và vị trí xung yếu cho những enzyme này cũng rất bất thường.

Tính chất đặc biệt của vị trí xung yếu (target site) có thể được tạo ra bằng cách phối hợp sự chuyên tính của methylase với sự chuyên tính của restriction endonuclease. Thí dụ *DpnI* cắt DNA tại sequence G^mA-T-C để tạo ra những đầu dây flush. Enzyme này tỏ ra không hiệu quả, nó cần phản ứng methyl hóa các gốc adenine trong cả hai dây DNA, để phản ứng phân cắt xảy ra. Vì vậy DNA của *dam*⁺ *E. coli* được cắt bằng *DpnI*, trong khi DNA của *dam*⁻ mutants không cắt (McClelland và ctv. 1984). Người ta đã phối hợp phản ứng chuyên tính của modification methylase M.Taq I hoặc M.Cla I với *DpnI*. Enzyme M.Taq I methyl hóa cả hai dây của chuỗi mã TCGA, để tạo ra T-C-G^mA. Từ đó, trong DNA không có phản ứng methyl hóa ở vị trí này nữa, mà bị phân cắt bởi *DpnI*, phản ứng xảy ra ở chuỗi mã có tính chất octanucleotide: T-C-G^mA- T-C-G-A. Tương tự như vậy, M.Cla I methyl hóa ở cả hai dây của chuỗi mã A-T-C-G-A-T để sản xuất ra chuỗi mã A-T-C-G^mA-T, rồi *DpnI* phân cắt ở chuỗi mã có tính decanucleotide ATCGATCGAT.

6.3.2. Điện di trên gel xung điện trường và đảo trường

Một loại hình mới về điện di được phát triển bởi Schwartz và Cantor (1984) để xử lý các phân tử DNA có độ dài từ 2000 kb trở lên. Kỹ thuật này sử dụng agarose và chất đệm bình thường như kỹ thuật cũ trong điện di. Nó chỉ cải tiến ở chỗ tạo xung điện định kỳ, điện trường theo hướng trục giao (orthogonal). Họ dùng YAC để chứng minh cho kỹ thuật mới này. Mỗi nhiễm thể nhân tạo là một mẫu DNA riêng, được phân tích Southern. Sự khác biệt nhau giữa các băng được phát hiện theo sự rối loạn dòng điện (perturbation) của hướng chạy DNA trên gel, và tùy theo mức độ kéo dài phân tử DNA. Thời gian bất hoạt của những phân tử này trong dung dịch tự do có chức năng siêu nhạy cảm của trọng lượng phân tử. Điện di trên xung điện trường trục giao PFGE (viết tắt từ chữ pulse-field gel electrophoresis) ngày càng được cải tiến nhiều hơn để phục vụ cho nghiên cứu genome (Carle và Olson 1984), nhưng khuyết điểm của nó là các mẫu DNA không chạy được theo đường thẳng.

Khó khăn này đã được giải quyết phần nào nhờ phát triển kỹ thuật điện di gel đảo trường FIGE (viết tắt từ chữ field-inversion gel electrophoresis) (Carle và ctv. 1986). Kỹ

thuật này khá tốt đối với mẫu DNA lớn hơn 2000 kb, mà không cần công cụ tạo trường trực giao. FIGE sử dụng máy điện di như phương pháp cũ, chỉ cải tiến điện trường tạo xung cả theo hai hướng kết hợp : tới (forward) và ngược lại (reverse), có thời gian nghỉ ngắn trong mỗi pha. Xung làm hình thành một time-ramp (tái khuếch đại có tính thời gian). DNA sẽ chạy theo một đường thẳng, và có thể so sánh trên gel, tạo ra sự điều tiết với những marker, và sự giải thích kết quả với phương pháp Southern. Nó đơn giản hơn kỹ thuật PFGE.

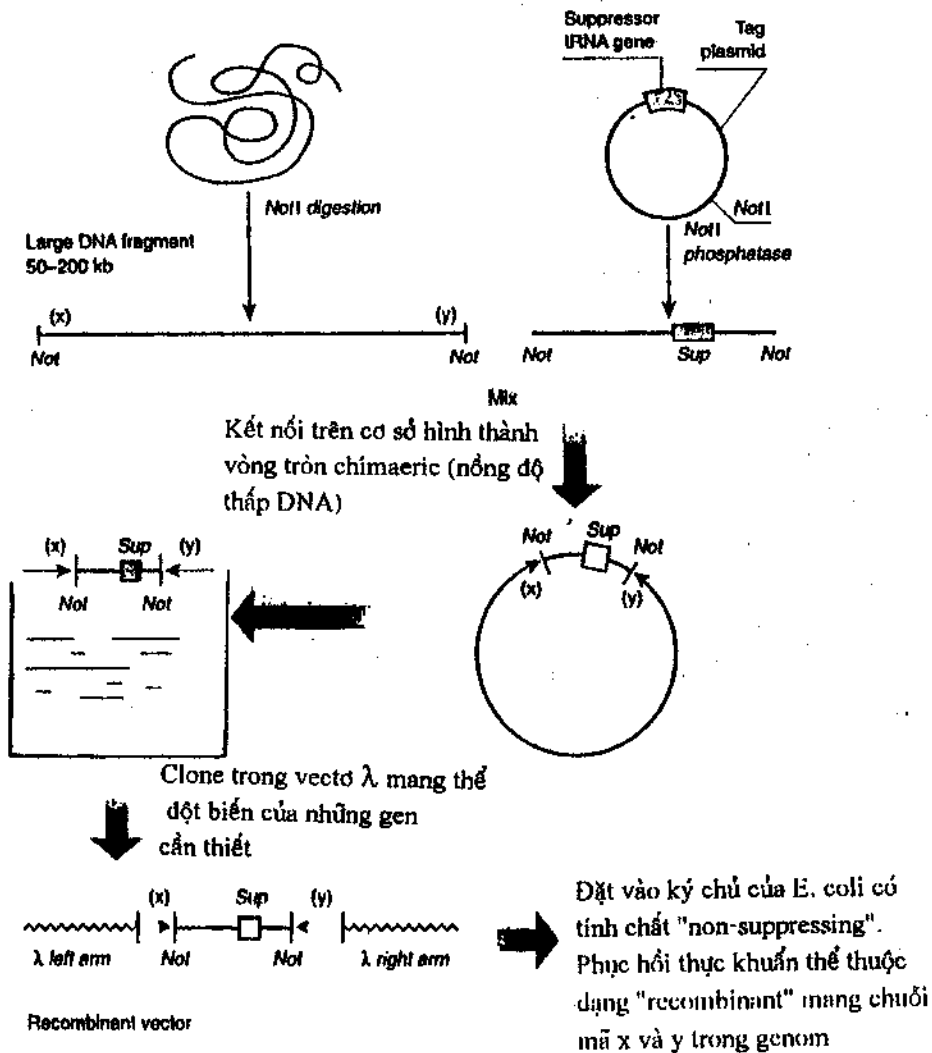
6.3.3. Kỹ thuật chromosome jumping

Kỹ thuật này được mô tả bởi Collins và Weissman (1984), Poustka và Lehrach (1986).

Nó tùy thuộc vào sự tuần hoàn của rất nhiều đoạn phân tử DNA, sau đó là cloning DNA từ vị trí bao phủ điểm nút (đóng mở) của vòng tuần hoàn này, làm cho các chuỗi mã DNA vốn định vị ở một khoảng cách khá xa nhau trong genome, được mang vào nằm gần nhau.

Những DNA được clone hóa tại điểm nút (closure site) sẽ hình thành một kho lưu trữ jumping. Kho lưu trữ có tính chất jumping như vậy cho phép tiến trình chromosome walking trên khoảng cách rất dài được thúc đẩy nhanh hơn. Hình 6-3: cho thấy phương pháp tạo ra một kho lưu trữ jumping sử dụng *NotI* làm men phân cắt genome DNA của người. Enzyme này chỉ cắt DNA của lớp động vật có vú. Phương pháp này được hỗ trợ bởi phân tích FIGE và Southern blotting các DNA của genome đã được tiêu hóa với *NotI*. Tuy nhiên, với một tiến trình tiêu hóa hoàn chỉnh, người ta phải thu nhận được một cách gián tiếp các đoạn phân tử *NotI* chồng lấp lên nhau. Giải pháp đặt ra là sử dụng các clone liên kết với nhau, có các đoạn phân tử được clone hóa theo phương pháp cổ điển, với một vị trí phân cắt của *NotI* ở bên trong (Poustka và ctv. 1987). Hình 6-3 cũng cho thấy hiện tượng tiêu hóa từng phần với một restriction endonuclease để tạo ra đoạn phân tử DNA lớn (Collins và Weissman 1984). Kỹ thuật này còn được áp dụng để jump 100 kb trong locus gen điều khiển u xơ ở nang (Collins và ctv. 1987) và jump 200 kb trong vùng gen kiểm soát tình trạng bệnh Huntington (Richards và ctv. 1988).

DNA phân tử cao



Hình 6.3 : Xây dựng kho lưu trữ theo kỹ thuật jumping (Old và Primose 1995)

6.4. KHO LƯU TRỮ cDNA

cDNA là phân tử DNA có tính bổ sung (complementary DNA) đối với RNA thông tin (mRNA), nó biểu hiện ra như một thành phần của một genome. Một bản sao phân tử DNA của mRNA được tổng hợp bằng enzyme reverse transcriptase. Nó chứa đựng thông tin di truyền của những gen đã được thể hiện trong genome. Phân tích chuỗi mã của các dòng vô tính cDNA được lấy ra một cách ngẫu nhiên từ kho lưu trữ cDNA sẽ giúp cho chúng ta gia tăng số chuỗi mã các gen đã được định tính trong thực vật và động vật, đồng thời cung cấp những kiến thức về sự chuyên tính của mô của những gen thể hiện.

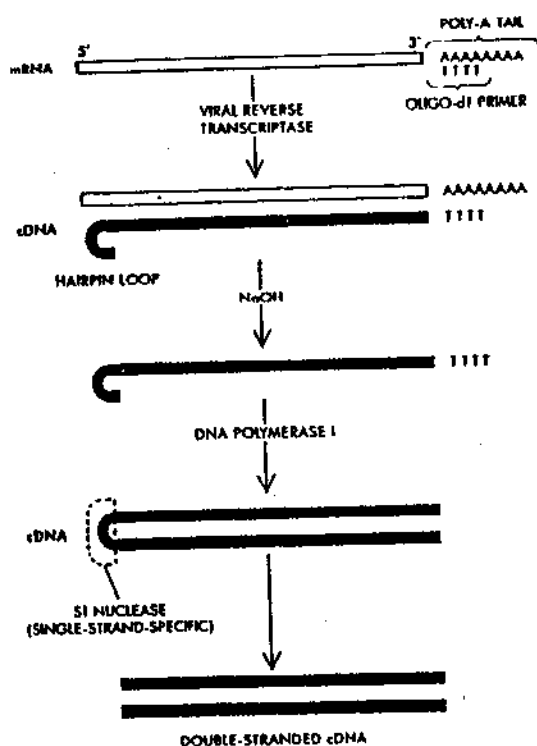
Genome của sinh vật bậc cao không có những chuỗi mã đơn giản của nucleotide với

dạng dây thẳng. Các exon và intron được phân biệt trong quá trình chuyển mã, và nhiều loại hình chuỗi mã có tính lặp lại (repetitive sequence) đã được công bố. Trong phân tích genome, việc phát hiện thể đa hình, xây dựng bản đồ liên kết gen, bản đồ vật lý đều cần được hoàn thiện bằng cách sử dụng các cDNA. Người ta tổng hợp cDNA từ mRNA và xây dựng tín hiệu di truyền (code) cho exon của genome. Theo mục tiêu này, cDNA là một công cụ quan trọng để làm rõ hơn cấu trúc của genome và phản ánh cách thể hiện của những gen từ các mô khác nhau (Sasaki và ctv. 1994). Những gen được thể hiện như vậy cho chúng ta nhận thấy đặc tính của sinh vật. Loài sinh vật này khác với loài sinh vật kia nhờ các gen có tính chất chuyên biệt cho loài. Điều này rất quan trọng trong phân loại (hình 6-4).

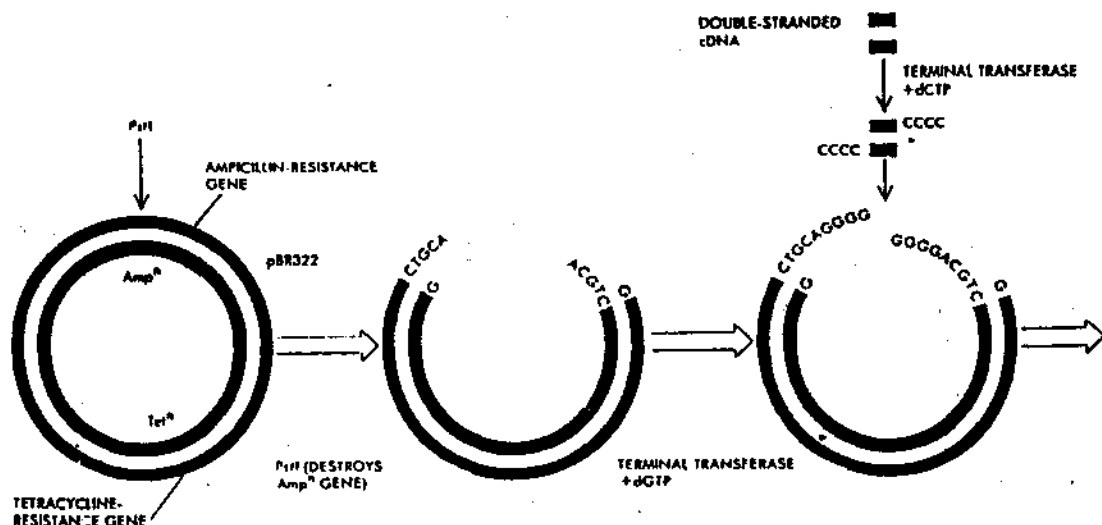
Phân tích đầu tiên về quần thể của các gen được thể hiện (expressed genes) trên mô của não bộ con người đã được công bố (Adams và ctv. 1991, 1992). "Gen được thể hiện" đã giữ các cDNA, chúng được đọc mã từng phần (dùng kỹ thuật sequencing), rồi được so sánh với các số liệu có trong Ngân Hàng Gen, áp dụng phép tính BLAST algorithm, rồi giải mã các sequence của amino acid tương phản với PIR database, áp dụng phép tính FASTA algorithm. Có khoảng 17% các clone đã được mã hóa đối với các gen đã được công bố của *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster* và nhiều loài thực vật. Nhóm nghiên cứu khác về genome con người có những công bố rõ ràng hơn về sự khác biệt của thể hiện gen trong các mô khác nhau, sử dụng kỹ thuật sequencing ngẫu nhiên ở đầu 3' của đoạn phân tử *MboI* trên cDNA (Ohkubo và ctv. 1992). Bên cạnh đó, nhiều công trình cDNA clone của *Arabidopsis thaliana* (Hofte và ctv. 1993), *Oryza sativa* (Uchimiya và ctv. 1992), *Zea mays* (Keith và ctv. 1993), *Brassica napus* (Park và ctv. 1993) đã được công bố về chuỗi mã. Đối với cây lúa, người ta đã tìm thấy khoảng 75 cDNA nucleotide sequences trong PIR release 32 (Sasaki và ctv. 1994). Nhiều tính trạng của cây lúa rất có giá trị, tuy nhiên có rất ít gen điều khiển những tính trạng này, được công bố về chuỗi mã. Do đó, việc phân lập các gen đặc biệt và làm rõ cấu trúc của gen rất cần thiết cho công tác lai tạo giống lúa đạt hiệu quả cao.

6.4.1. Nghiên cứu chuyên đề trên cây lúa (case study)

Kích thước genome cây lúa ước khoảng $4,3 \times 10^8$ bp trên mỗi đơn bội thể (Arumuganathan và Earle 1991). Kích thước như vậy là nhỏ nhất trong các loài cây trồng và rau hoa. Kích thước genome của *Arabidopsis thaliana* nhỏ hơn cây lúa 1/3 và có ít hơn các tính trạng nông học có giá trị. Do đó cây lúa là mô hình kiểu mẫu trong thực vật để nghiên cứu về sinh học phân tử. Phân tích đọc chuỗi mã từng phần của những clone cDNA cây lúa và sử dụng chúng như những RFLP marker, nhằm xây dựng bản đồ liên kết gen có chất lượng cao là mục tiêu của Rice Genome Program thuộc NIAR, Nhật Bản trong mấy năm gần đây (Sasaki và ctv. 1994). Từ bản đồ này, người ta sẽ phát triển bản đồ vật lý để đáng hơn. Người ta còn lập một catalog cDNA của cây lúa, nó sẽ rất có ích trong việc so sánh các gen được thể hiện với nhau.



Hình 6.4a : Tổng hợp cDNA dây đôi từ mRNA. Chuỗi oligo dT gắn được lại với đuôi poly A của mRNA. Đoạn ligo-T phục vụ như một primer cho phản ứng reverse transcriptase, dùng mRNA làm dây template để tổng hợp cDNA dây đơn. Kết quả là cuối dây cDNA tạo thành móc hình cong. Khi mRNA được xử lý với NaOH, móc cong này trở thành primer cho DNA polymerase I, hoàn thành việc cặp đôi dây DNA. Móc cong này được cắt ra bởi S1 nuclease, cho ra phân tử cDNA dây đôi.



Hình 6.4b : Tổng hợp cDNA bằng đuôi homopolymeric. Gen Amp^r của pBR322 được cắt bởi PstI. Terminal transferase và dGTP được dùng để thêm vào đuôi poly G tại đầu 3'. Terminal transferase và dCTP được dùng để gắn vào đuôi poly C tại đầu 3' của cDNA dây đôi. Đoạn cDNA này được đưa vào trong plasmid. Người ta cho enzym vào trám các lỗ trống trên hai dây của plasmid. Kết quả cho ra một plasmid có tính chất recombinant, để rồi nó được gắn vào vi khuẩn. Người ta dùng nó để xét nghiệm vi khuẩn kháng tetracycline, không kháng ampicillin

6.4.1.1. Đặc tính của kho lưu trữ cDNA

Sasaki và ctv. (1994) đã lấy ra 3138 clone và phân tích được 2778 clone có chiều dài gắn vào (insert length) là 0,3 kb. Kích thước của những đoạn phân tử cDNA dùng cho việc kết nối (ligation) chỉ cần khoảng 0,7 kb. Kích thước có tính chất thoải mái nhất (abundant) của cDNA cho phép gắn vào là 0,6 - 0,9 kb (giá trị trung bình là 0,88 kb, SD = 0,36 kb). Điều này cho thấy tính chất abundance của những cDNA đối với kích thước như vậy trong kho lưu trữ, là tính chất có thể sử dụng được một cách hiệu quả.

6.4.1.2. Phát sinh ra những sequence tag được thể hiện

2778 clone có khả năng insert trên 0,3 kb đã được công bố chuỗi mã từ đầu 5'.

172 clone cũng được công bố chuỗi mã từ đầu 3'.

Trong những clone được công bố chuỗi mã, có 2491 clone đạt yêu cầu cho phân tích FASTA (Sasaki và ctv. 1994). Trong một công trình khác, người ta đã công bố chuỗi mã 7200 clone, trong đó 5250 từ callus và 1946 từ rễ lúa bằng kỹ thuật single run partial sequencing, rồi lưu trữ 4231 chuỗi mã trong DNA Databank của Nhật (Kurata và ctv. 1994). Nếu những nucleotide sequence chồng lấp trên 100 base ở vùng có tính chất tương đồng, và sự khác biệt của những chuỗi mã này lớn hơn 10%, thì nó được xem như các thể đồng dạng (isoforms) của một gen. Sau khi kiểm tra tính chất redundancy (phong phú, thừa thãi) của những chuỗi mã này, 2259 EST được tạo ra và bổ sung số liệu vào DNA Databank. EST là thuật ngữ được viết tắt từ chữ "expressed sequence tags" (thể đánh dấu chuỗi mã đã được thể hiện). Những EST này phát sinh từ 2159 clone độc lập và có 111 chuỗi mã phát xuất từ đầu 3' (Sasaki và ctv. 1994).

6.4.1.3. Những chuỗi mã được thể hiện

Trong một thí nghiệm, người ta đã sử dụng 1500 cDNA clone từ callus và 1500 clone từ rễ lúa để phân tích RFLP (Kurata và ctv. 1994). Theo kết quả phân tích Southern với những clone này, các cDNA có khả năng copy thấp và có tính đồng nhất dưới hình thức RFLP, được dùng làm những thể thăm dò cho phân tích quần thể phân ly F2. Một vài nhóm gen có tính chất multi-copy và các gen mã hóa isozyme cũng được ghi vào bản đồ. Theo kết quả phân tích sequence của 7200 clone từ callus, rễ lúa và thư mục cDNA của các mô khác, có hơn 80 loại gen protein khác nhau (ribosome), và 61 clone giống với 15 gen histone (Kurata và ctv. 1994). Trong số này, chỉ có một vài clone có thể được dùng làm RFLP phục vụ cho công tác lập bản đồ. Đó là những băng biểu hiện sự kiện lai đặc biệt, bên cạnh những băng bình thường. Phân tích RFLP bằng cách sử dụng các clone cDNA sẽ cho chúng ta biết mức độ redundancy của gen trong genome cây lúa, và mức độ đa hình của những chuỗi mã kề cận (flanking), hay chuỗi mã có tính can thiệp (intervening) của gen giữa giống lúa trồng japonica và indica.

6.4.2. Tính chất phong phú (abundance) của mRNA

cDNA phát xuất từ mRNA, do đó nó không có chuỗi mã của intron. Người ta rất hiếm khi phân lập được cDNA từ những mRNA thuần khiết (Old và Primrose 1995). Thông thường một kho lưu trữ cDNA được chuẩn bị và được thanh lọc theo kiểu sequencing từng phần. Trước khi triển khai các bước tiếp theo, người ta cần xem xét bản chất của quần thể mRNA trong những mô tế bào. Trong nhiều mô và tế bào, mRNA có mặt một cách rộng khắp, với mức độ phong phú (abundance) rất khác nhau; có nghĩa là, một vài mRNA có mặt với số lượng lớn trong tế bào, những loại mRNA khác chỉ có chừng một vài copy trong tế bào.

Bảng 6.1 : Xếp hạng tính phong phú của những quần thể mRNA điển hình

Nguồn	Số lượng mRNA khác nhau	Mức phong phú (mol / tế bào)
Tế bào chất của gan chuột	9	12000
poly (A)	700	300
	11500	15
Polysoma ống dẫn trứng của gà	1	100000
poly (A)	7	4000
	12500	5

Bảng 6.1 là một ví dụ minh chứng tính phong phú của mRNA. Ống dẫn trứng của gà có mRNA được xếp vào loại superabundant. Đó là mRNA mã hóa ovalbumin, protein chính trong lòng trắng trứng. Do đó quần thể mRNA này sẽ được sử dụng tạo ra ovalbumin cDNA, không có vấn đề gì khó khăn cho việc thanh lọc. Những clone này có thể được xác định bằng cách thanh lọc một số nhỏ các recombinant. Tiến trình giải mã sẽ diễn ra tốt đẹp.

Một phương pháp khác để có cDNA phong phú là clone các cDNA trực tiếp trong vectơ M13 thí dụ như M13 mp8. Một nhóm clone có thể được phân tích sequence ngay sau đó và xác định trên cơ sở polypeptide của mỗi một chuỗi mã. Phương pháp được áp dụng khá thành công có thuật ngữ là "shotgun sequencing" do Putney và ctv. (1983) thực hiện, nhằm xác định chuỗi mã DNA của 178 cDNA recombinant của mô cơ bắp người ta. Chuỗi mã hoàn chỉnh của amino acid đã đáp ứng cho 19 protein có quan hệ với mô cơ bắp này.

Đối với các dòng cDNA có độ phong phú thấp, nó rất có ích cho việc lập kho lưu trữ cDNA. Theo nguyên tắc, người ta có 10^5 clone là đủ trong trường hợp mRNA có độ phong phú thấp, ở hầu hết các loại tế bào. Một lần nữa, vectơ thực khuẩn thể lambda tỏ ra có hiệu

quả trong việc mang số lượng lớn các cDNA clones. Những vectơ như λ gt10, λ NM1149, λ ZAP hoặc λ gt11 đều rất phù hợp cho kỹ thuật cloning cDNA.

Trước khi cloning, chúng ta có cần nhiều mRNA hay không? Chỉ trong những trường hợp hết sức đặc biệt mà sự thuần khiết (purification) đã sẵn sàng rồi, công việc thu thập nhiều mRNA mới được thực hiện tốt đẹp. Kỹ thuật được sử dụng phổ biến là phân đoạn kích thước mRNA thông qua điện di trên gel (Pennica và ctv. 1983). Một kỹ thuật khác để tạo ra cDNA là sử dụng các primer cố tính chọn lọc rất cao, với qui trình PCR cũng rất chọn lọc.

6.4.3. Kỹ thuật cloning cDNA có chiều dài nguyên vẹn

Chúng ta có hai cDNA, mà sự tổng hợp dây đơn số hai của nó có tính chất self-primed (tự tạo ra phản ứng với primer trong quá trình annealing). Hiện tượng này sẽ hình thành một cDNA dây đôi, xoắn chéo, có vòng dạng hair pin. Đoạn vòng thắt eo này sẽ bị loại ra ngay sau đó bằng men nuclease S1. Tiến trình này làm mất đi một số chuỗi mã thích ứng với đầu 5' của mRNA, làm tổn hại đến dây đôi cDNA. Vì lý do đó, kỹ thuật self-priming trong tổng hợp dây số hai rất ít được thực hiện. Người ta cố gắng tìm cách cải tiến phương pháp này, khắc phục những khó khăn nói trên.

Một trong những cách làm đơn giản nhất do Land và ctv. (1981) thực hiện là hình thành đuôi dC của dây đơn cDNA, theo sau đó là phản ứng priming bằng oligo (dG) để tổng hợp dây thứ hai, hiện tượng vòng hair pin không xảy ra, không cần xử lý nuclease S1, và kết quả cuối cùng là tạo ra dây cDNA có chiều dài nguyên vẹn.

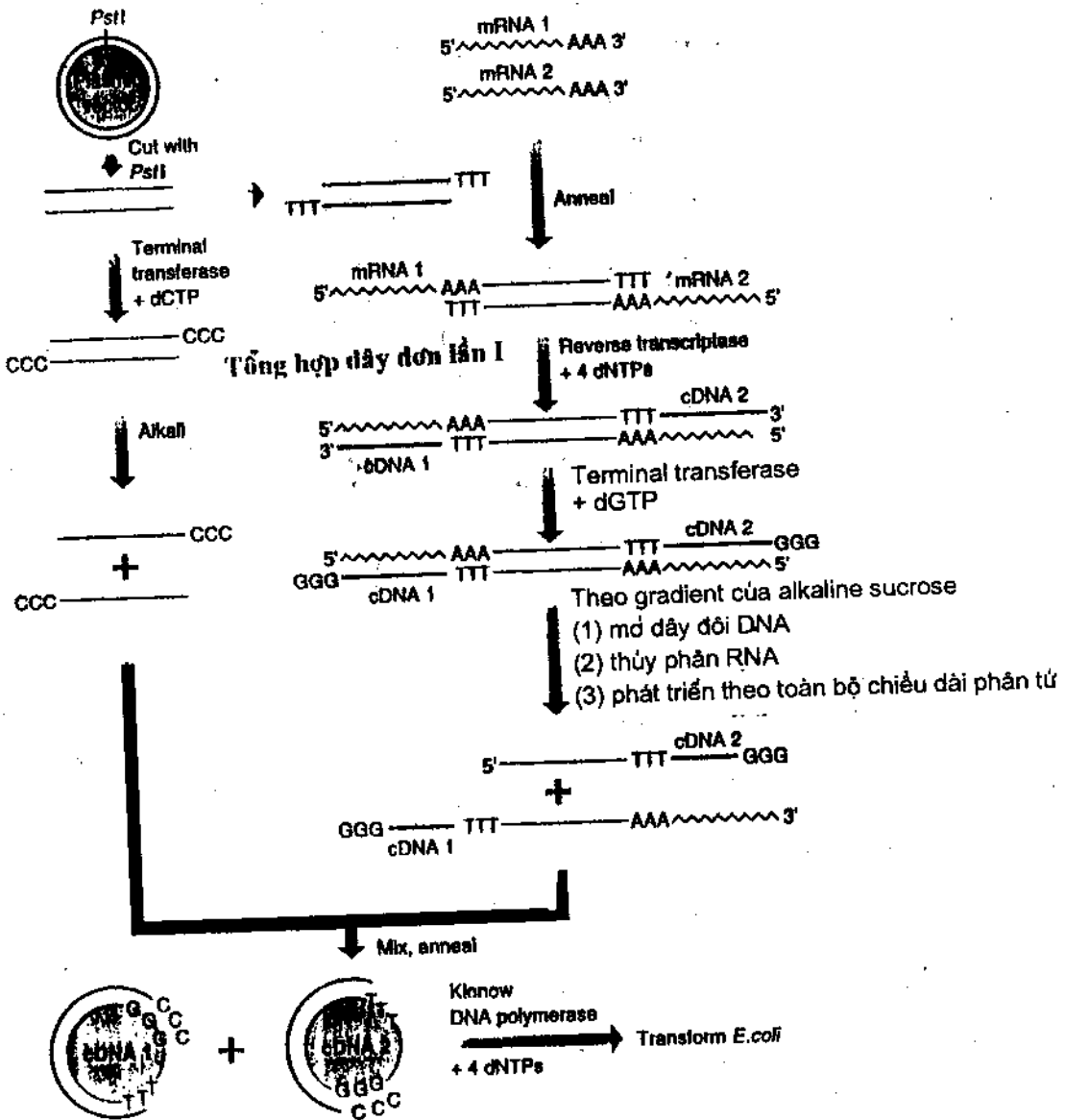
Hai phương pháp cải tiến của Okayama và Berg (1982), Heidecker và Messing (1983) cũng cho hiệu quả rất cao, nhằm hạn chế tối đa việc sử dụng nuclease S1 (hình 6-5). Sự chuyển mã ngược có chiều dài nguyên vẹn đã được ghi nhận. Phân tử có tính chất "lai" giữa RNA-DNA tạo ra kết quả sinh tổng hợp dây thứ nhất, dây này là cơ chất của phản ứng men terminal transferase. Một cDNA như vậy sẽ không kéo dài đến đầu mút của mRNA. Dòng cDNA này sẽ có gốc 3-hydroxyl có tính chất đệm, làm cho nó trở thành một cơ chất rất kém khả năng tạo đuôi (tailing).

Lợi ích của phương pháp Okayama và Berg là cDNA có thể được gắn vào vectơ có định hướng rõ ràng. Đó là một dạng ứng dụng để gắn cDNA vào các vectơ thực khuẩn thể T3, T7 hoặc SP6 promoter. Bởi vì một copy của RNA có thể được tổng hợp *in vitro* từ một dây cDNA có trước với RNA polymerase của thực khuẩn thể được tinh khiết hóa (Old và Primrose 1995).

6.4.4. Sử dụng các primer ngẫu nhiên trong kỹ thuật cloning cDNA

Chúng ta đã thảo luận nhiều về sự tổng hợp cDNA được bắt đầu bằng phản ứng tại đuôi poly(A)⁺ của mRNA, bằng phản ứng priming với các chuỗi mã oligo(dT). Như vậy, chúng ta sẽ có ba hạn chế trong phương pháp này:

- Không phải tất cả các RNA đều mang chuỗi mã poly(A) có 3'-terminal. Người ta có thể thêm vào chuỗi mã poly(A) *in vitro* với poly(A) polymerase thuần khiết, nhưng điều này có thể gây ra nhiều vấn đề bất lợi.
- Phân tử mRNA có kích thước lớn có thể khó xử lý, bởi vì người ta không hi vọng sẽ tổng hợp và chạy sequence các clone có quan hệ với primer dạng oligo(dT).



Tổng hợp dây đơn lần II

Hình 6.5 : Kỹ thuật cloning cDNA có chiều dài nguyên vẹn (Okayama và Berg 1982, Heidecker và Messing 1983)



Hình 6.6 : Tổng hợp cDNA nhờ công cụ tổng hợp nhanh có tên TimeSaver, với nhiều primer khác nhau. Cột 1 là mRNA tiêu chuẩn với Olydo (dT)12 18 primer [0,5 μ g]. Cột 2, cùng một RNA, nhưng dùng primer có hai chức năng [5 μ g]. Cột 3, globin mRNA với primer thông dụng [0,375 μ g]. Cột 4, mRNA chuẩn với pb(N), primer [0,074 μ g]. Cột 5, cùng một RNA và primer như cột 4, nhưng cho nhiều primer hơn [7,4 μ g]

- Sai lệch ở đầu 3' : phản ứng priming xảy ra tại đầu 3' của poly(A)⁺ mRNA, sinh tổng hợp cDNA thường không hoàn toàn, cho nên thư mục cDNA thường rất giàu chuỗi mã 3'-terminal.

Hai giới hạn đầu tiên rất phổ biến trong kỹ thuật cloning các RNA của genome từ RNA virus. Còn giới hạn thứ ba có thể trở nên quan trọng ở những kho lưu trữ cDNA được làm ra trong vectơ : chẳng hạn như λ gt11, λ ZAP, hoặc trong kho lưu trữ cDNA được thanh lọc trên cơ sở sự thể hiện các polypeptide dung hợp (fusion polypeptides). Những hạn chế như vậy có thể được khắc phục bằng cách áp dụng phương pháp priming để sinh tổng hợp dây thứ nhất cDNA, không có oligo(dT), mà là những primer có tính chất ngẫu nhiên,

thuộc dạng oligonucleotide. Thông thường primer có một sự trộn chung của tất cả những hexadeoxynucleotide có tính chất tổng hợp về mặt hóa học. Những thể lai này tại các vị trí ngẫu nhiên dọc theo phân tử RNA, sự tổng hợp cDNA, ngân ấy sẽ phát sinh chuỗi mã cDNA có kích thước vừa đủ cho kỹ thuật cloning.

6.4.5. PCR & kho lưu trữ cDNA và kho lưu trữ dna của genome

Với nhiều thành tựu diệu kỳ của PCR (phản ứng chuỗi polymerase) trong những năm gần đây, người ta có xu hướng thực hiện các kho lưu trữ DNA của genome và kho lưu trữ cDNA trên cơ sở PCR còn gọi là "PCR-based approach". Nó cho chúng ta một phương pháp nhanh hơn, đơn giản hơn trong quá trình xây dựng kho lưu trữ và thanh lọc kho lưu trữ. Những nội dung này sẽ được thảo luận trong các phần sau.

TÀI LIỆU THAM KHẢO

1. Adams MD, JM Kelley, JD Gocayne. 1991. *Complementary DNA sequencing: Expressed sequence tags and human genome project*. Science 252:1651-1656
2. Adams MD, M Dubnick, AR Kerlavage. 1992. *Sequence identification of 2,375 human brain genes*. Nature 355:632-634
3. Arumuganathan K, ED Earle. 1991. *Nuclear DNA content of some important plant species*. Plant Mol. Biol. Rep. 9:208-218
4. Bailey CR, MJ Butler, ID Normansell, RT Rolands, DJ Winstanley. 1984. *Cloning of Streptomyces clavuligerus genetic locus involved in clavulanic acid biosynthesis*. Biotechnology 2:808-811
5. Bender W, P Spierer, DS Hogness. 1983. *Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in Drosophila melanogaster*. J. Mol. Biol. 168:17-33
6. Bibb MJ, MJ Bibb, JM Ward, SN Cogen. 1985. *Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to Streptomyces*. Mol.Gen. Genet. 199:26-36
7. Blackburn E. 1985. *Artificial chromosomes in yeast*. Trends in Genetics 1:8-12
8. Carle GR, MV Olson. 1984. *Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis*. Nucleic Acids Res. 12:5647-5664
9. Clarke L, J Carbon. 1980. *Isolation of yeast centromere and construction of functional small circular chromosome*. Nature 287:504-509
10. Collins FS, SW Weissman. 1984. *Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method*. Proc. Natl. Acad. Sci. USA 81:6812-6816
11. Dhaese P, C Hussey, M van Montagu. 1984. *Thermo-inducible gene expression in Bacillus subtilis using transcriptional regulatory elements from temperate phage Φ 105*. Gene 32:181-194
12. Ehrlich SD, B Niaudet, B Michel. 1982. *Use of plasmids from Staphylococcus aureus for cloning of DNA in Bacillus subtilis*. Current Topics in Microbiol. And Immunol. 96:19-29
13. Gray O, S Chang. 1981. *Molecular cloning and expression of Bacillus licheniformis Lactamase Gene in Escherichia coli and Bacillus subtilis*. J. of Bacteriol. 145:422-428
14. Hanahan D, M Meselson. 1980. *Plasmid screening at high colony density*. Gene 10: 63-67

15. Heidecker G, J Messing. 1983. *Sequence analysis of Zein cDNAs by an efficient mRNA cloning method*. Nucleic Acids Res. 11: 4891-4906
16. Hofte H, T Desprez, J Amselem. 1993. *A inventory of 1152 expressed sequence tags tained by partial sequencing of cDNAs from Arabidopsis thaliana*. Plant J. 4:1051-1061
17. Keith CS, DO Hoang, BM Barrett, B Feigelman, MC Nelson, H Thai, C Baysdorfer. 1993. *Partial sequence analysis of 130 randomly selected maize cDNA clones*. Plant Physiol.. 101:329-332
18. Kerem BS, JM Rommens, JA Buchanan, D Markiewicz, TK Cox, A Chakravati, M Buchwald, LC Tsui. 1989. *Identification of the cystic fibrosis gene: genetic analysis*. Science 245:1073-1080
19. Kurata N, Y Nagamura, K Yamamoto, Y harushima, N Sue, J Wu, BA Antonio, A Shomura, T Shimizu, SY Lin, T Inoue, A Fukuda, T Shimano, Y Kuboki, T Toyama, Y Miyamoto, T Kiriara, K Hayasaka, A Miyao, L Monna, HS Zhong, Y Tamura, ZX Wang, T Momma, Y Umehara, M Yano, T Sasaki, Y Min e. 1994. *A 300 kilobase interval genetic map of rice including 883 expressed sequences*. Nature Genetics 8:365-372
20. Land H, M Grey, H Hanser, W Lindermaier, G Schutz. 1981. *5'-terminal sequences of eukaryotic mRNA can be cloned with a high efficiency*. Nucleic Acids Res. 9: 2251-2266
21. Leach DRF, F Stahl. 1983. *Viability of lambda phages carrying a perfect palindrome in the absence of recombination nucleases*. Nature 305:448-451
22. Maniatis T, RC Hardison, E Lacy, J Lauer, C O'Connell, D Quon, GK Sim, A Efstratiadis. 1978. *The isolation of structural genes from libraries of eukaryotic DNA*. Cell 15:687-701
23. McClelland M, LG Kessler, M Bittner. 1984. *Site-specific cleavage of DNA at 8- and 10-base pair sequences*. Proc. Natl. Acad. Sci. USA 81:983-987
24. Murray AW, JW Szostak. 1983. *Construction of artificial chromosomes in yeast*. Nature 305:189-193
25. Ohkubo K, N Hori, R Mabota, T Niiyama, A Fukushima, Y Kojima, K Matsubara. 1992. *Large scale cDNA sequencing for analysis quantitative and quantitative aspects of gene expression*. Nature Genet. 2:173-179.
26. Okayama H, P Berg. 1982. *High-efficiency cloning of full length cDNA*. Mol. Cell. Biol. 2:161-170
27. Old RW, SB Primrose. 1995. *Principles of Gene Manipulation*. NG CARR, Univ. of Warwick. 6th Ed. London, UK.

28. Park YS, JM Kwak, OY Kwon, YS Kim, DS Lee, MJ Cho, HH Lee, HG Nam. 1993. *Generation of expressed sequence tags of random root clones of Brassica napus by single-run partial sequencing*. Plant Physiol. 103:359-370
29. Pennica D, WE Holmes WJ Kohr. 1983. *Cloning and expression of human tissue-type plasminogen activator cDNA in E. coli*. Nature 301:214-221
30. Poustka A, H Lehrach. 1986. *Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics*. Trends Genetics 2:174-179
31. Putney SD, WC Herlihy, P Schimmel. 1983. *A new troponin T and cDNA clones for 12 different muscle proteins, found by shotgun sequencing*. Nature 302:718-721
32. Richards JE, TC Gilliam, JL Cole, ML Drumm, JL Wasmuth, JF Gusella, FS Collins. 1988. *Chromosome jumping from D4S10 (G8) toward the Huntington disease gene*. Proc. Natl. Acad. Sci. USA 85:6437-6441
33. Riordan JR, JM Rommens, BS Karem, N Alon, R Rozmahel, G Zbyszko, J Zielenski, S Lok, N Plavsic, JL Chou, ML Drumm, MC Iannuzzi, FS Collins, LC Tsui. 1989. *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA*. Science 245:1066-1073
34. Rommens JM, Iannuzzi, BS Karem, ML Drumm, G Melmer, M Dean, R Rozmahel, JL Cole, D Kennedy, N Hidaka, M Zsiga, M Buchwald, JR Riordan, LC Tsui, FS Collins. 1989. *Identification of the cystic fibrosis gene: chromosome walking and jumping*. Science 245:1059-1065
35. Sasaki T, J Song, Y Koga-Ban, E Matsui, F Fang, H Higo, H Nagasaki, M Hori, M Miya, E Murayama, T Takiguchi, A Takasuga, T Niki, K Ishimaru, H Ikeda, Y Yamamoto, Y Mukai, I Ohta, N Miyadera, I Havukkala, Y Min e. 1994. *Toward cataloguing all rice genes: large-scale sequencing of randomly chosen rice cDNAs from a callus cDNA library*. The Plant J. 6(4):615-624
36. Schwartz DC, CR Cantor. 1984. *Separation of yeast chromosomal-size DNAs by pulsed gel gradient gel electrophoresis*. Cell 37:67-75
37. Uchimiya H, S Kidou, T Shimazaki. 1992. *Random sequencing of cDNA libraries reveals a variety of expressed genes in cultured cells of rice (Oryza sativa L)*. Plant J. 2:1005-1009
38. William DM, RG Schoner, EJ Duvall, LH Preis, PS Lovett. 1981. *Expression of Escherichia coli trp gene and the mouse dihydrofolate reductase gene cloned in Bacillus subtilis*. Gene 16:199-206
39. Yansura DG, DJ Henner. 1984. *Use of the Escherichia coli lac repressor and operator to control gene expression in Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 81:439-443.

BẢN ĐỒ DI TRUYỀN & BẢN ĐỒ VẬT LÝ

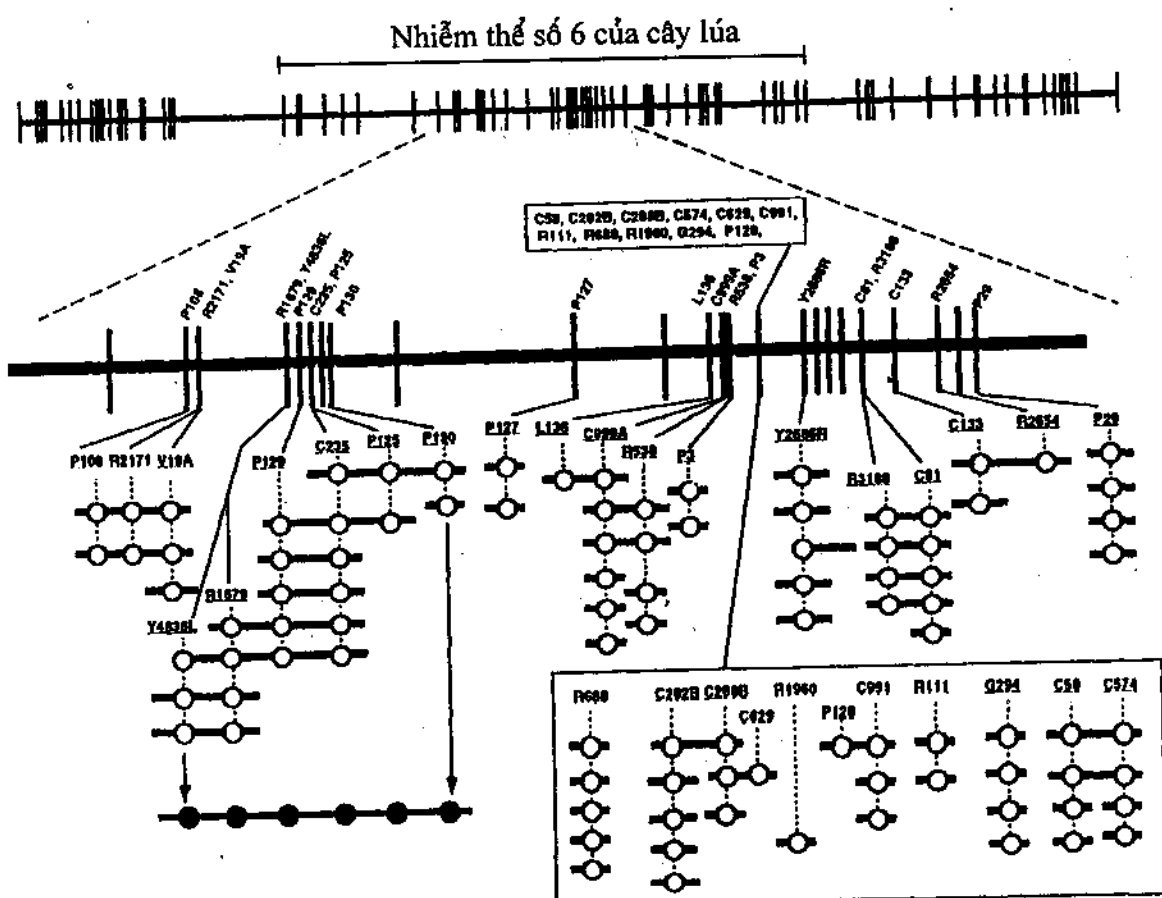
7.1. KHÁI NIỆM BẢN ĐỒ TRONG DI TRUYỀN CỔ ĐIỂN

Người ta đã dùng hiện tượng tái tổ hợp, hay trao đổi chéo, để định vị gen trên nhiễm thể, trong di truyền cổ điển (conventional). Thí dụ có tính chất giáo trình đã được thực hiện trên *Neurospora*. Dựa trên hiện tượng trao đổi chéo trong giảm phân, người ta quan sát sự tương quan giữa trao đổi chéo di truyền học và hình chữ thập tế bào học dưới kính hiển vi. Sự ổn định của tần số tiền giảm và hậu giảm của đôi alen dấu tính nói riêng và các đôi alen khác nói chung, cho phép người ta đặt giả thiết là locus của các gen cố định trên nhiễm thể. Khoảng cách giữa hai gen cùng nằm trên nhiễm thể được tính bằng centi Morgan (cM). Bản đồ lập theo phương pháp này được gọi là bản đồ gen trao đổi chéo, hay tái tổ hợp (recombination).

7.2. BẢN ĐỒ LIÊN KẾT GEN VÀ MARKER PHÂN TỬ

Kỹ thuật được áp dụng có thuật ngữ là "genome mapping" cho phép chúng ta nghiên cứu các quá trình có tính chất hình thái, sinh lý, phát triển trong đó có những biến dị di truyền xảy ra. Genome mapping yêu cầu một hướng dẫn tối thiểu có tính chất tiên đoán. Để khai thác hiệu quả PCR, người ta sử dụng YAC (nhiễm thể nhân tạo làm từ men) và phân tích có tính chất so sánh giữa các loài thực vật có quan hệ xa, làm nền tảng cho sự hiểu biết về sinh thực vật và canh tác cây trồng (thí dụ ở hình 7-1 trên nhiễm thể số 6 của cây lúa).

Genome mapping là một khái niệm có tính chất tổng hợp từ di truyền cổ điển, nó là công cụ của **sinh học phân tử** và **toán sinh học**. Genome mapping là một phương pháp rất có hiệu quả trong nghiên cứu sinh học thực vật, nó lấy nguyên tắc phân tích liên kết gen của di truyền cổ điển làm nền tảng, cộng thêm phương tiện hiện đại của công nghệ sinh học để mapping các gen có tính chất giản đơn hoặc các gen có tính chất phức tạp về kiểu hình. Genome mapping phù hợp trên nhiều loài thực vật, bởi vì nó yêu cầu thời gian nghiên cứu không quá dài, dễ sử dụng trên quần thể khá lớn. Việc sưu tập các thể biến dị (variants) về hình thái, sinh lý, và phát triển đều có thể thực hiện, cung cấp nguồn gen dồi dào cho nghiên cứu. Genome mapping cho chúng ta phương tiện trực tiếp nghiên cứu số gen ảnh hưởng đến một tính trạng nào đó, vị trí của những gen này trên nhiễm thể, ảnh hưởng biến dị có tính chất số lượng của những gen này (Paterson 1996). Mapping sẽ thực hiện trên từng gen đơn và các gen điều khiển tính trạng số lượng (QTL) với phương pháp thực hiện rất đa dạng.



Hình 7.1 : Một thí dụ về cách sắp xếp theo đường thẳng của YAC clone, dọc theo bản đồ liên kết gen của nhiễm thể số 6 trên cây lúa. Các thanh ngang có hình tròn là những YAC clone. Từ Y4836L đến P130, chín YAC clone chồng lấp lên nhau hình thành một contig, được biểu thị bằng mũi tên đen. Mười một DNA marker đồng phân ly tại một locus, biểu thị bằng một khung hình chữ nhật nằm trên bản đồ liên kết gen, được lai với hơn 20 YAC clone mà vẫn không đủ cho phản ứng chồng lấp. Khoảng cách vật lý tương ứng với locus di truyền, được ước lượng khoảng > 1Mb (Sasaki và ctv. 1996).

Việc tạo dòng trên cơ sở bản đồ gen [map-based cloning] tạo điều kiện thuận lợi cho công tác phân lập các gen quan trọng, với các thông tin tối thiểu có tính chất tiên đoán. Việc tạo dòng trên cơ sở bản đồ gen của các loài thực vật thượng đẳng rất phức tạp do genome có qui mô quá lớn, các đoạn DNA quan trọng có tính chất lặp lại, và đa bội thể. Do đó nhiều gen quan trọng đã được cloned ở thực vật như cây *Arabidopsis* bằng đột biến gắn bên trong [insertional mutagenesis] hoặc bằng lai tạo cơ chất [subtractive hybridization] (Felman và ctv. 1989, Sun và ctv. 1992). Đối với cây trồng, người ta cố gắng phát triển càng nhiều càng tốt các bản đồ RFLP và thư mục các YAC đối với các gen có quan hệ với những tính trạng quan trọng về kinh tế.

Bản đồ di truyền chi tiết trên cơ sở RFLP đã được thực hiện trên nhiều loài thực vật (Paterson và ctv 1992). Nổi tiếng nhất là bản đồ gen của cà chua với 1400 marker trên 12 nhiễm thể (Tanksley và ctv 1992), của *Arapidopsis* với 500 marker trên 5 nhiễm thể (Chang và ctv. 198, Nam và ctv 1989, Retter và ctv. 1992). Gần đây có nhiều công trình được công bố trên cây thông, lúa mạch, đậu phộng, cải dầu, kiều mạch, lúa mì, mía, bông vải, v.v...(Devey và ctv. 1991, Graner và ctv. 1991, Kochert và ctv. 1991, Song và ctv. 1991, Wang và ctv. 1991, Anderson và ctv. 1992).

Sự phối hợp lại để có một bản đồ di truyền có thể được đánh giá bằng phương pháp thống kê (Hulbert và ctv. 1988), nhưng người ta cũng có những trắc nghiệm sinh học để xem xét. Mapping thể đa hình trên đoạn phụ cuối nhiễm thể với 162 bp cho thấy bản đồ di truyền của 4 nhiễm thể cà chua trong quãng 5-10 cM (Broun và ctv 1992). Có một sự tương phản rõ nét trong những kết quả gần đây là lai tạo được *in situ* các dòng có số copy ít, được ghi trên bản đồ nhiễm thể cây lúa, cho thấy bản đồ di truyền đều nằm ở đoạn cuối của các nhiễm thể. Đó là một mâu thuẫn chưa giải quyết (Gustafson và Dille 1992).

Hầu hết các bản đồ di truyền đầu tiên của cây trồng đều được thực hiện bằng phương pháp chọn lọc gia phả, đối với mức độ đa dạng cao về DNA marker của nó. Người ta còn thấy rằng, việc lai xa tạo điều kiện cho genetic mapping vì phần lớn các nguồn gen của những cây trồng quan trọng rất hạn chế.

Kỹ thuật PCR-based genotyping rất có triển vọng trong việc lập bản đồ di truyền. (Paterson và Wing 1993). Hạn chế của phương pháp này là muốn phát triển các STMs [sequence-tagged microsatellites] người ta phải có một sự đầu tư rất lớn về kỹ thuật sequencing. Các phương tiện được đề nghị như RAPD, PCR và arbitrary primer PCR (primer ngẫu nhiên) sẽ làm giảm tối thiểu chi phí sử dụng các primer của chuỗi mã có tính chất quasi-arbitrary trong quá trình khuếch đại DNA. Một bản đồ liên kết gen trên cơ sở RAPD đã được công bố (Retter và ctv 1992). Tính trội của tất cả RAPD marker được thể hiện trong F2 (Allard 1959). Mức độ tin cậy của RAPD cũng chưa đủ so với yêu cầu làm bản đồ liên kết gen, mặc dù các arbitrary primer PCR marker tương đối đáng tin cậy (Paterson và Wing 1993). Hạn chế của RAPD trong việc lập bản đồ di truyền đầu tiên sẽ còn tồn tại, nhưng người ta nhấn mạnh rằng PCR có sử dụng arbitrary primer vẫn có giá trị trong việc làm phong phú vùng nhiễm thể chưa xác định đối với các DNA marker.

Các quần thể được sử dụng trong kỹ thuật làm bản đồ di truyền thực vật thường là:

- RIL (recombinant inbred lines) : các dòng cận giao tái tổ hợp, với điều kiện bố mẹ phải thật sự tương phản nhau về tính trạng mình muốn lập bản đồ, cá thể F2 phải đủ lớn (>200) và lấy ngẫu nhiên, từ đó chọn dòng cận giao liên tục.
- NIL (nearly isogenic lines): các dòng gần như đẳng gen / trong trường hợp nghiên cứu gen kháng bệnh trên cây trồng, mà pathogen (nấm, vi khuẩn có quá nhiều nỗi phức tạp).

- F2 và F3, bố mẹ P1, P2, quần thể hồi giao BC1 (F1 x P1) và BC2 (F1 x P2).
- Quần thể đơn bội kép (double haploid) viết tắt là DH.

7.2.1. Mapping có tính chất so sánh

Nhiều chuỗi mã di truyền đã được bảo quản thông qua rất nhiều loài sinh vật. Trong khi đó các chuỗi không mã hóa lại có mức độ đa dạng cao hơn. Việc lập bản đồ di truyền của các DNA đang được bảo quản (như cDNAs) cho phép chúng ta nghiên cứu thứ tự của các gen trên nhiễm thể ở các loài có nguồn gốc xa nhau. Thông qua mapping có tính chất so sánh, cùng với tính chất tương đương của marker trên nhiều loài cây khác nhau [synteny], các điểm breakpoint đối với những thứ bậc gen khác nhau, đã được minh chứng trên cà chua, khoai tây, cao lương và bắp (Tanksley và ctv 1992, Bonierbale và ctv. 1988, Gebhart và ctv 1991, Hulbert và ctv. 1990). Mapping có tính chất so sánh có những tác dụng như sau :

- Tạo ra một vài sự kiện đáng lưu ý có tính chất tiến hóa giữa các loài.
- Ứng dụng bản đồ gen của loài này trên loài khác (lúa, lúa mì, lúa mạch, cao lương).
- Giải thích hiện tượng phức tạp do đa bội thể.

7.2.2. Mapping kiểu hình và đánh dấu gen

Mapping kiểu hình là một thuyết minh đầu tiên trong việc thiết lập bản đồ di truyền của thực vật. Các DNA marker có liên kết chặt sẽ tạo điều kiện thuận lợi cho công tác chọn lọc gián tiếp các tính trạng khó đo lường. Những marker này tạo điều kiện ban đầu cho việc cloning các gen điều khiển các tính trạng quan trọng. Các DNA có liên kết chặt chẽ với những tính trạng di truyền có thể được xác định bằng RAPD PCR trên quần thể các dòng gần như đẳng gen [NFL] (Martin và ctv. 1991). Gần đây, kỹ thuật này đã được phát triển mạnh ở bất cứ vùng nhiễm thể nào đối với DNA marker, sử dụng các DNA pool có tính chất tổng hợp, với nhiều marker mới được phân lập.

Phương pháp DNA pooling làm tăng khả năng phân lập các QTL marker (Michelmore và ctv. 1991). Nó có giá trị trong việc mapping các QTL và du nhập [introgression] QTL vào các genotype khác nhau. (QTL: viết tắt từ chữ quantitative trait loci, có nghĩa là các loci của những tính trạng số lượng)

7.2.3. Mapping các tính trạng số lượng

Hầu hết các tính trạng nông học quan trọng, bao gồm năng suất, phẩm chất dinh dưỡng và tính chống chịu môi trường, đều là tính trạng di truyền số lượng (Allard 1960, Hallauer và Miranda 1988). Khả năng xử lý các gen này là một yêu cầu, đối với việc cải tiến giống cây trồng có tính chất ổn định. Trong 10 năm qua, đã có nhiều báo cáo về việc sử dụng phương pháp marker phân tử để khám phá, lập bản đồ gen, và định tính những loci

điều khiển tính trạng di truyền số lượng của cây trồng (Paterson và ctv. 1988, Keim và ctv. 1990, Stuber và ctv. 1992, Fatokun và ctv. 1992, Anderson và ctv. 1993, Hayes và ctv. 1993, Wang và ctv. 1994). Mặc dù có những thành công nhất định, nhưng có rất ít trường hợp, trong đó kỹ thuật marker phân tử đã tạo ra giống cây trồng mới, với sự cải tiến của một hoặc nhiều tính trạng số lượng nào đó (hình 7-2).

Sự khám phá QTL và sự phát triển giống hiện nay là hai quá trình tách rời nhau. Giai đoạn đầu là khám phá QTL. Các dòng bố mẹ được xác định, có những tính trạng số lượng đối nghịch nhau [thí dụ như năng suất, phẩm chất]. Người ta lai vật liệu bố và mẹ, để có quần thể con lai phân ly, trong đó marker đã được sử dụng để xác định những QTL được liên kết. Giai đoạn thứ hai là sử dụng kiến thức định vị bản đồ QTL để sáng tạo ra giống cây mới.

7.2.3.1. Định nghĩa QTL

- Tính trạng số lượng [QT]: là tính trạng thể hiện sự phân bố chuẩn, liên tục trong quần thể lớn và chưa qua một chọn lọc nào.
- Kiểu hình [P] : sự đo lường hiệu quả của các cá thể.
- Ảnh hưởng kiểu gen [G]: ảnh hưởng của yếu tố di truyền, kiểm soát kiểu hình, được đo lường thông qua độ lệch với giá trị trung bình quần thể [μ].
- Ảnh hưởng môi trường [e]: ảnh hưởng của môi trường trên quần thể.
- Ảnh hưởng tính cộng [a] : một phần của ảnh hưởng môi trường.
- Độ lệch tính trội [d] : ảnh hưởng kiểu gen không có sự tham gia của [a], ảnh hưởng tương tác giữa hai alen trong cùng locus.
- Epistasis [i] : ảnh hưởng tương tác giữa hai, ba, hoặc nhiều loci

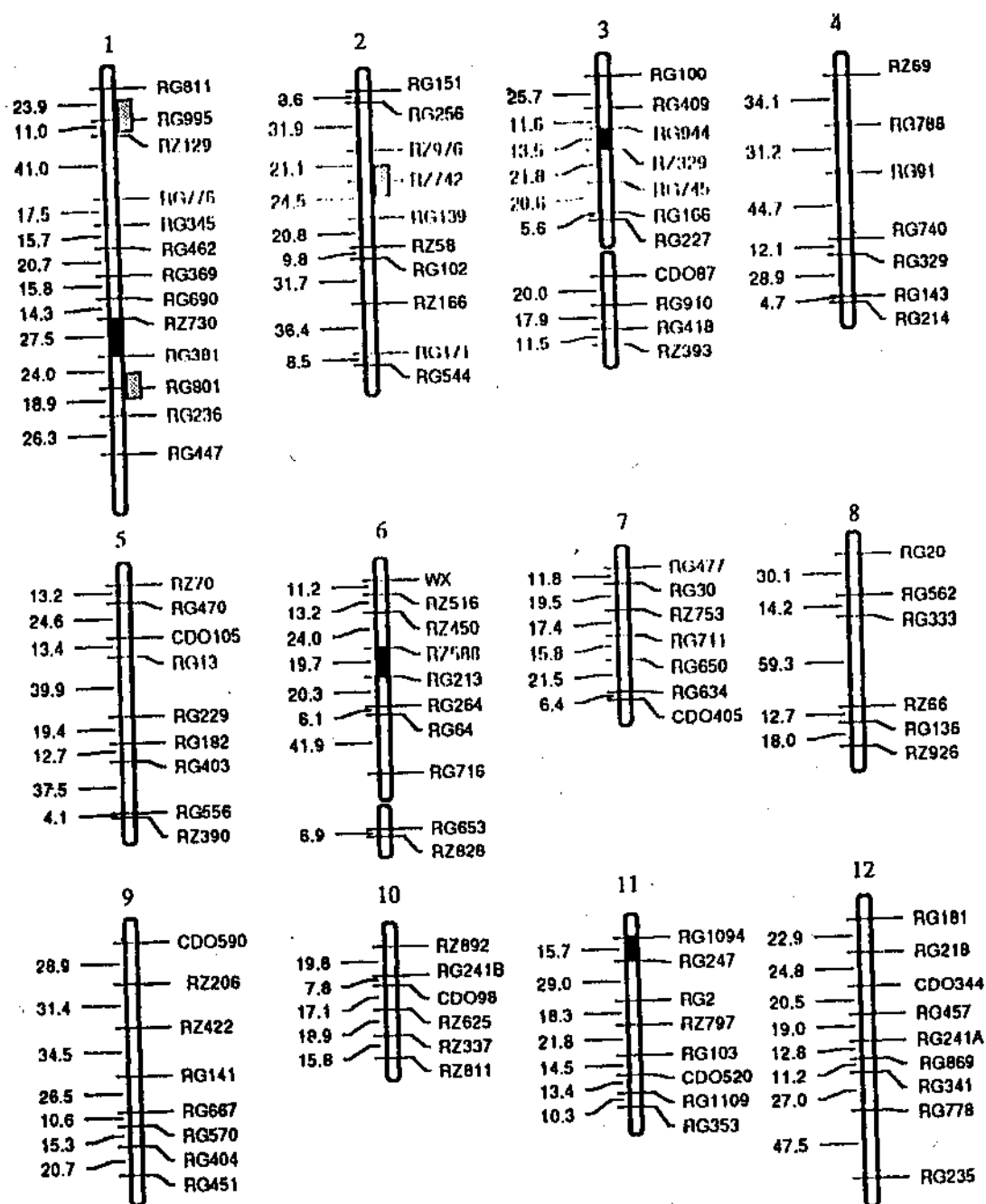
$$P = \mu + G + e = \mu + a + d + i + e$$

- Sai lệch trong liên kết [Linkage disequilibrium] : các loci không thể kết hợp ngẫu nhiên để tạo ra sự tăng, hoặc giảm kiểu gen nào đó, do sự liên kết quá chặt hoặc yếu tố khác, so với trường hợp nó không liên kết. Mức độ sai lệch trong liên kết được xác định bằng mức độ liên kết, sự chọn lọc, v.v....

7.2.3.2. Cơ sở di truyền về QTL

Thí dụ trong cặp lai: MMTT x mmtt

Trong đó M/m là hai alen của locus có tính chất marker, và T/t là hai alen của QTL. Chúng ta giả định rằng : có một giá trị liên kết r giữa M và T trên nhiễm thể mà chúng hiện diện. Trong F₂, nếu r = 0.5, thì sẽ có một tái tổ hợp tự do [free recombination]. Tất cả kiểu gen của F₂ đều có thể xuất hiện theo tỉ lệ mong muốn.



Hình 7.2 : Bản đồ RFLP của cây lúa có từ một quần thể phân ly bao gồm 231 cá thể của tổ hợp lai Palawan/IR42. Khoảng cách được tính theo đơn vị centiMorgan, nhờ MAPMARKER, version 1.0 [Macintosh]. Các vùng đậm đen của nhiễm thể 1, 3, 6 và 11 là những đoạn chứa QTL. Những khung dính bên phải nhiễm thể là các marker loci, có ảnh hưởng tương tác không alen giữa RG801 và hai loci khác. Chữ ở bên phải của nhiễm thể là tên marker, ở bên trái là khoảng cách di truyền tính bằng cM.

Nếu $r = 0$ thì $MMTt = 0,25$

$MmTt = 0,5$

$mmTt = 0,25$

Do đó, sự phân loại các alen có tính chất marker sẽ tạo ra kết quả phân loại các QTL alen, và hiệu số giữa những giá trị trung bình các lớp sẽ là:

$$Mm - mm = 2a(1 - 2r)$$

Khả năng khám phá sự phối hợp QTL / marker tùy thuộc vào cả hai giá trị r và a .

7.2.3.3. Sử dụng quần thể trong phân tích phối hợp QTL / marker

Bất cứ cấu trúc của quần thể nào cũng có thể được sử dụng để phân tích phối hợp QTL / marker. Loại quần thể thường được sử dụng là:

1. Cây F2 của tổ hợp lai giữa hai đồng hợp tử
2. Dòng F3 từ những cá thể F2
3. Dòng cận giao tái tổ hợp [RIL]
4. BC1
5. Dòng đơn bội kép [DH]

Loại hình khác như: tổ hợp lai/quần thể cũng có thể được sử dụng. Người ta sử dụng lai diallel để lập bản đồ gen có tính chất ưu thế lai [dị hợp tử]. Lai thử nghiệm [test cross] cũng được khuyến cáo. Con lai từ những cặp lai của một quần thể [diallel] cho đến con lai đã ổn định sau quá trình cận giao đã được phân tích.

Qui mô của quần thể phải đủ lớn. Chúng ta sẽ gặp phải những trở ngại sau đây: (a) quần thể càng lớn, giá thành càng đắt, thí dụ quần thể DH, (b) chi phí để tính toán giá trị kiểu hình, (c) chi phí để tính toán giá trị kiểu gen [RFLP]. Theo các nghiên cứu gần đây, người ta thường sử dụng quần thể có qui mô từ 150 đến 300 cá thể. Tuy nhiên, Stubber và ctv. (1992) đã cho rằng: qui mô của quần thể nhỏ sẽ mang lại một kết quả không chính xác, và không đủ điều kiện để xác định QTL. Tiêu chuẩn của IRRI hiện nay là 200-300 cá thể của F2 và quần thể RI. Muốn có một bản đồ marker tốt, và tránh được hiện tượng liên kết giả (pseudo linkage), chúng ta phải một quần thể không sai lệch quá đáng so với hai bố mẹ.

7.2.3.4. Phương pháp tính để phát hiện sự phối hợp marker / QTL

Phân tích tuyến tính dạng đơn:

$$P = \mu + aX + e$$

Trong đó hàm số P là giá trị kiểu hình của một cá thể, μ là giá trị trung bình quần thể, và thừa số e là ảnh hưởng môi trường với $N(0, e)$, X là biến số thể hiện liều lượng các

alen; đóng góp vào (0, 1, 2, đối với F2 ; 0, 2 đối với quần thể RI ; và 0, 1 đối với BC1).

Khi chúng ta chưa xác định liều lượng các alen đóng góp vào, chúng ta có thể sử dụng marker alleles kế bên làm kiểm chứng. Hệ số gốc phương trình a là kết quả tính toán sự đóng góp của biến số X.

Nếu tần suất tái tổ hợp $r < 0.5$ và ảnh hưởng tính cộng của gen đủ lớn, thì hệ số gốc a sẽ không được là zero hoặc gần bằng không. Do đó, trắc nghiệm giả thuyết đối nghịch, với $a = 0$, được xem như là kiểm chứng sự có mặt, hoặc vắng mặt của QTL ở trên, hoặc ở xung quanh marker. Hệ số gốc a là phép đo sự tham gia của QTL. Khi có sự kết hợp giữa QTL và marker, ảnh hưởng tính cộng sẽ được ước đoán. Mức độ ước đoán này tùy thuộc vào tần suất tái tổ hợp "r".

ANOVA một chiều :

Phối hợp QTL / marker có thể được trắc nghiệm bằng cách phân tích phương sai một chiều [one-way ANOVA]. Người ta trắc nghiệm hiệu số khác biệt của những giá trị trung bình của kiểu gen

$$MM = a(1-2r) + 2r(1-r)d$$

$$Mm = d(1-2r+2r^2)$$

$$mm = -a(1-2r) + 2r(1-r)d$$

$$\text{nếu } r = 0.5, \text{ thì } MM - mm = 0$$

$$r = 0, \text{ thì } MM - mm = 2a$$

$$0 < r < 0.5 \text{ thì } MM - mm = 2a(1-2r)$$

Do đó phép thử F để $MM = mm$ sẽ cho chúng ta một xác định về sự liên kết có ý nghĩa. Tuy nhiên, chúng ta không thể phân biệt những ảnh hưởng do liên kết chặt chẽ, với ảnh hưởng do kiểu gen quá mạnh.

Chỉ có hai lớp trong quần thể DH, RI. Do đó, nó chỉ chứa đựng ảnh hưởng có tính cộng mà thôi. Nhưng quần thể F2 có ba lớp marker, với hai độ tự do; chúng ta phải tách ra ảnh hưởng có tính cộng [thuộc dạng tuyến tính], ảnh hưởng có tính trội [thuộc dạng quadratic], mỗi ảnh hưởng chiếm một độ tự do. Để làm được việc này, người ta cần phải lập ra một loạt các giá trị tương phản có tính trực giao [orthogonal], mỗi cái có một độ tự do. Sau đó, người ta phân tích phương sai để hoàn tất những biến số có tính kiểm định này. Phép thử F sẽ giúp chúng ta biết được mức độ có ý nghĩa ấy. Phân tích này còn được gọi là phân tích theo xu hướng [trend analysis].

Phân tích trên cơ sở tính trạng [Trait-based analysis] :

Lebowitz và ctv. (1987) sáng tác ra phương pháp này, trong đó cá thể nào có tính

trạng ưu việt đều được chọn lọc. Tần suất marker được so sánh với tần suất giả định. Những marker có sự khác biệt về tần suất giữa cái chọn lọc và cái ước đoán như vậy, sẽ được giả định là có kết hợp với QTL, gây ảnh hưởng trên tính trạng nghiên cứu, do một giả thuyết có thuật ngữ là "genetic hitchhiking". Nếu quần thể sai lệch quá nhiều so với ước đoán, người ta sẽ sử dụng một mẫu ngẫu nhiên làm đối chứng. Trắc nghiệm Z được sử dụng như sau:

$$Z = \frac{p_1 - p_2}{\left(\frac{p_1 q_1}{2n_1} + \frac{p_2 q_2}{2n_2} \right)^{1/2}}$$

trong đó p_1 và p_2 / q_1 và q_2 , là tần suất alen của quần thể có chọn lọc và quần thể ngẫu nhiên, theo thứ tự. n_1 và n_2 là qui mô của quần thể. Z được xem như phân bố chuẩn.

Ưu điểm của phương pháp này là qui mô hữu hiệu của quần thể được mở rộng. Nhược điểm của nó là chỉ có một tính trạng được phân tích trong cùng một thời gian. Với chi phí rất đắt của RFLP hiện nay, điều này thật sự không kinh tế lắm.

Lập bản đồ cách quãng QTL :

Phương pháp phân tích trên cơ sở tính trạng gặp phải vấn đề : (a) nếu $r > 0$ ảnh hưởng của gen có thể được dự đoán, (b) nó không dự đoán được vị trí của QTL. Lander và Botstein (1989) đã sáng tạo ra phương pháp được gọi là lập bản đồ cách quãng [interval mapping] với kỹ thuật mô phỏng [maximum likelihood]. Không những chỉ có điểm marker được phân tích, mà còn khoảng chứa marker cũng được tìm thấy. Một lô LOD được ghi nhận với một đỉnh cho thấy hầu hết vị trí của một QTL. Các tác giả đã soạn thảo một chương trình puter riêng cho phân tích này (hình 7-3).

Mặc dù phương pháp phân tích cách quãng có thể tốt hơn phương pháp tuyến tính đơn, hoặc phương pháp ANOVA một chiều, nhưng người ta đã chứng minh rằng cả hai đều cho ra cùng một kết quả (Stuber và ctv. 1992, Nodari và ctv. 1993).

Kiểu phương trình dự đoán có tính đa dạng [multiple predictive model] :

Kiểu phương trình dự đoán có tính đa dạng được sử dụng phổ biến, để tạo ra các model, nhằm ước đoán ảnh hưởng của những marker loci cùng một lúc. Nó được thực hiện giống như phương pháp tuyến tính đơn. Kiểu gen của RFLP marker được chuyển thành biến kiểm chứng, và trở thành các biến độc lập. Tiếp theo đó, hàm thuận và nghịch, hằng số, hệ số gốc được hoàn chỉnh. Với một tiêu chuẩn đặt ra từ trước, người ta có thể xác định các loci, tại đó nó giải thích hầu hết các biến dị di truyền của quần thể. Do vấn đề liên kết gen, người ta phải quan sát các marker loci, xem nó có liên kết rất chặt với nhau hay không, bởi vì nó có thể xuất hiện trong cùng một QTL. Phép thử đồng tuyến [colinearity test] có thể được sử dụng.

Người ta đã cải tiến phương pháp này bằng cách sử dụng những marker phối hợp với QTL trong phân tích tuyến tính đơn (Cowen và ctv. 1992). Trong trường hợp như vậy, một bộ khóa các marker chủ lực được cài trong mô hình. Những marker không có ý nghĩa có thể cho thêm vào mô hình để xem xét mức độ ý nghĩa có tăng lên theo giá trị R^2 hay không.

Tính toán epistasis :

Tương tác giữa hai QTL có thể được phân tích thông qua ANOVA hai chiều, hệ phương trình đa tuyến, và phân tích cách quãng, với mô hình mở rộng một cách đơn giản của phương thức một marker - tại một thời gian. Cowen và ctv. (1992) đã xác định bốn QTLs, có tương tác với nhau, chiếm 47% biến dị, trong các giá trị trung bình gia đình, xét theo khả năng của bao phần sản xuất ra các cấu trúc phối giống nhau.

7.3. BẢN ĐỒ VẬT LÝ

Làm thế nào người ta có thể chuyển đổi "khoảng cách di truyền" đo bằng tái tổ hợp thành "khoảng cách vật lý" đo bằng số cặp base giữa hai marker di truyền ?

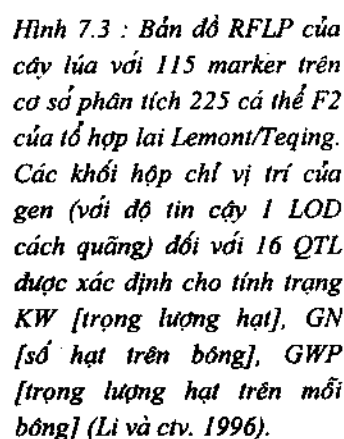
Bảng 7-1 cho chúng ta thấy: số lượng tổng cộng DNA của một genome trong thực vật bậc cao so với genome của người và vi khuẩn. Những dự đoán này làm cơ sở cho việc xem xét mối tương quan giữa khoảng cách di truyền và khoảng cách vật lý.

Trong khi sử dụng một bản đồ di truyền hoàn chỉnh để ước đoán chiều dài di truyền tổng cộng của một genome, người ta có thể tính số lượng trung bình của DNA tương ứng với khoảng cách di truyền của tái tổ hợp 1% (nghĩa là 1 cM). Số lượng có tính chất vật lý của DNA tương ứng với 1cM thay đổi tùy theo loài thực vật bậc cao, ước khoảng 280 kb trong *Arabidopsis* cho đến 7000 kb trong lúa mạch (có thể nhiều lần hơn nữa trong hoa tulip).

Bảng 7.1 : Kích thước genome của thực vật, người và vi khuẩn (ước đoán)

	Amt. DNA/nhân (pg)	AAmt. DNA/cM	#150 kb clones bao phủ 5x
Vi khuẩn	0,002 - 0,01	N.A	127
Người	6,0	1000 kb	92.100
Thực vật			
<i>Arabidopsis thaliana</i>	0,3	280 kb	4.600
Lúa	0,9	300 kb	13.800
Cao lương	1,3	400 kb	19.950
Caà chua	1,9	750 kb	29.160
Bông vải	4,5	400 kb	69.100
Lúa mạch (<i>H. vulgare</i>)	10,0	7000 kb	154.000
Tulip	6,4	N.A.	982.400

1 picogram (pg) = $0,965 \times 10^9$ bp



Thành tựu về phương pháp điện di trên xung điện trường, và vectơ YAC cho chúng ta một nguồn khá lớn về genome thực vật được tạo dòng [cloned] (Swartz và Cantor 1984, Burke và ctv. 1987).

7.3.1. Chuẩn bị megabase DNA

Phương pháp đáng tin cậy để phân lập DNA tương đối còn nguyên lành, chưa bị phân cắt là điều kiện đầu tiên trong thực hiện điện di xung điện trường và YAC cloning. Kỹ thuật phân lập megabase DNA đã được áp dụng trên nhiều loài cây trồng như đậu nành (Honeycutt và ctv. 1992), lúa (Sral và ctv. 1990), lúa mì, lúa mạch và kiều mạch (Cheung và Gale 1990).

7.3.2. Khoảng cách di truyền cổ điển và phân tử

Khi một marker phân tử được xác định có liên kết với một gen nào đó, người ta tin rằng sẽ thiết lập được mối quan hệ giữa khoảng cách di truyền và khoảng cách vật lý [tính theo phân tử] tại vùng xung quanh của gen này. Thông tin như vậy sẽ giúp cho chúng ta xác định tính chất walking hay jumping của một nhiễm thể. Mối quan hệ trung bình giữa khoảng cách di truyền và vật lý của một genome được tính toán theo kích cỡ genome và độ dài của bản đồ di truyền. Tuy nhiên, các giá trị tính toán hiện nay cho thấy, đối với bất cứ vị trí đặc biệt nào cũng đều có biến động rất lớn so với giá trị trung bình. Trong cà chua 1 cM tương đương 900 kb, nhưng ở tại vùng *Tm2a* của nhiễm thể số 9 thì 1 cM là 4-16 Mb (Ganal và ctv. 1989). Khoảng cách vật lý này được xem như là tiềm năng của sự tạo dòng trên cơ sở bản đồ. Ở lúa mì, tỉ lệ kb : cM lớn hơn rất nhiều so với cà chua, 1 cM tương đương 1 Mb ở vùng có chứa gen α -amylase trên nhiễm thể số 6 (Cheung và Gale 1990).

7.3.3. Kho lưu trữ YAC

Kho lưu trữ YAC đã được thành lập trên cây *Arabidopsis*, cà chua, bắp, lúa. Ở cà chua, Martin và ctv. (1992) ghi nhận có 22.000 clone, một nửa được tuyển chọn làm kho lưu trữ với các RFLP marker có liên kết chặt với 2 gen kháng bệnh *Tm2a* và *Pto*. Năm YAC đã được xác định trong thanh lọc, làm điểm khởi đầu cho sự walking của nhiễm thể đối với những gen này. Gần đây, kho lưu trữ YAC của cây bắp với 3 bộ chuẩn genome đơn bội thể đã được công bố (Edward và ctv 1992).

7.3.4. Bản đồ vật lý hoàn chỉnh

Bản đồ vật lý rất có giá trị trong nghiên cứu về tổ chức và sự tiến hóa của nhiễm thể cây trồng. Lại *in situ* cho chúng ta biết tổng quát về tổ chức genome. Các bản đồ "contig"- sử dụng RFLP map có mật độ cao - cosmid và YAC, cho chúng ta biết cụ thể trên từng genome. Một bản đồ cosmid có tính chồng lấp, có chứa 90-95% genome của *Arabidopsis*, đã được thành lập (Hauge và ctv. 1991). Có 125 YACs được xác định, với RFLPs thông qua 5 nhóm liên kết gen. Những YAC này đại diện cho 30 % genome của *Arabidopsis*.

7.3.5. Xác định gen

Phương pháp có tính chất trực tiếp nhất đối với việc xác định gen từ một YAC tiêu chuẩn nào đó hay một cosmid là bổ sung thêm một mutant do sự chuyển nạp. Giraudat và ctv (1992) đã chứng minh tính chất bổ sung của mutant ở gen *abi 3*, trong khi sử dụng một subclone của một trong ba cosmid có tính chồng lấp liên kết với RFLP. Một YAC hay một cosmid có thể được dùng như một chất thăm dò để xác định cDNA đặc biệt đối với những vùng nghiên cứu. cDNA bình thường hay đột biến đều có thể bị mã hóa trong các chuỗi nhằm phát hiện các đột biến tiềm ẩn. cDNA tốt nhất trong trường hợp này được dùng làm chất bổ sung có tính đột biến, hay sáng tạo ra đột biến trong kỹ thuật antisense. Arondel và ctv. (1992) đã xác định một bộ YAC có tính chồng lấp, phủ trên locus *fad3* của *Arabidopsis*. Người ta dùng một YAC để thanh lọc một thư mục cDNA làm từ hạt cây cải dầu (*Brassica napus*). Người ta cũng đã phân lập cDNA dị hợp và cho thấy tính chất bổ sung của alen *fad3*.

Khi chúng ta không tìm thấy cDNA nào, chúng ta phải đưa YAC vào như subclone đối với các đoạn có tính chồng lấp, và cố gắng sao cho việc bổ sung này trên từng subclone phải có tính độc lập. Nó đòi hỏi nhiều thời gian và tiền của, công sức

7.3.6. Contig

Chữ contig hay contiguous sequence diễn tả một đoạn dài của chuỗi mã được xây dựng bởi một số các sequence ngắn, sequence này có các đoạn phân tử chồng lấp lên nhau. Trong kỹ thuật YAC và BAC, người ta sáng tạo ra nhiều contig để thuận lợi cho việc nghiên cứu genome, đồng thời thuận lợi cho kỹ thuật hình thành bản đồ vật lý.

Trước đây, người ta chỉ dựa vào bản đồ liên kết gen với RFLP marker, sau đó clone hóa để hình thành bản đồ vật lý. Công việc này rất cồng kềnh mất nhiều thời gian và tiền của. Người ta tìm cách sử dụng STS marker, hoặc những marker khác trên cơ sở PCR, chuyển hóa chúng thành những clone, dùng YAC hoặc BAC vectơ, hình thành hàng loạt những contig, nhằm xây dựng bản đồ vật lý có chất lượng cao. Điều này sẽ cho phép các nhà chọn giống cây trồng khai thác thoải mái nguồn vật liệu trong ngân hàng gen, đẩy mạnh việc sử dụng nguồn gen địa phương, vốn rất hạn chế trong các chương trình lai cổ điển (hai, ba hoặc bốn bố mẹ trong mỗi lần lai tạo). Nhờ kết quả cloning và bản đồ vật lý, chúng ta có thể khai thác một lúc nhiều bố mẹ trong chương trình lai của mình. Thêm vào đó, với kỹ thuật MAS (chọn lọc giống nhờ marker phân tử), nhà chọn giống có thể rút ngắn thời gian và tiết kiệm tiền của, nhất là đối với các tính trạng khó thanh lọc bằng phương tiện thông thường: như lúa chịu lạnh, khô hạn, tạo dòng bất dục đực miễn cảm với nhiệt độ lạnh (TGMS) (Lang và ctv. 1997).

Nguyên tắc walking dọc trên nhiễm thể để xác định hàng loạt các dòng DNA (clone) có tính chất kề cạnh nhau - contigs - làm cơ sở cho một khối lượng nghiên cứu genome to lớn sau này.

- Kỹ thuật chromosome walking trong một vùng đặc biệt nào đó sẽ cung cấp một phương tiện để phân lập những gen, chúng được ghi nhận rằng: đã có mặt trong vùng đó, nhờ bản đồ di truyền (hay bản đồ liên kết gen và marker phân tử) được hình thành từ trước.
- Việc hình thành đầy đủ các contig trên những nhiễm thể, sẽ cung cấp nguồn tư liệu có ích trong tương lai, giúp chúng ta giảm đi những nghiên cứu có tính chất lặt vặt tại những vị trí chuyên biệt nào đó.

Việc xem xét đầu tiên là sự kết hợp lại các contig có qui mô cỡ nào trong mỗi bước của tiến trình, sao cho số lượng phân tử DNA gắn trong những cloning vectơ có thể được dùng để thành lập thư mục. Việc xem xét này giống như con dao hai lưỡi - tiến trình càng dài bao nhiêu thì thành tựu càng tốt hơn bấy nhiêu trong kết hợp lại những contig, nhưng bên cạnh đó, giải pháp đạt được sẽ ít hơn, bởi vì những gen mục tiêu (target gene) cần phải được xác định trên một đoạn phân tử DNA lớn hơn. Mỗi một nỗ lực như vậy trong kỹ thuật chromosome walking cần dùng thực khuẩn thể lambda làm vectơ, chỉ mang được 10.000 - 20.000 nucleotide (10-20 kb) các DNA ngoại sinh.

Cosmid có thể mang 35 kb DNA, và di chuyển đến mục tiêu nhanh hơn lambda (Paterson 1996). Do đó cosmid được áp dụng phổ biến để cải tiến quá trình vốn rất chậm hiện nay trong mapping có tính chất vật lý, trên động vật và thực vật, đặc biệt trong trường hợp sinh vật có kích thước genome quá lớn như người, và một số cây trồng.

Sự phát triển YAC đã làm thay đổi cách nghĩ trước đây về bản đồ vật lý, nhất là lĩnh vực thực hành các contig trên nhiễm thể của sinh vật bậc cao thuộc eukaryote. Những thư mục đầu tiên của YAC đã có những clone, mà khả năng gắn của nó là 150 kb. Hiện nay, nó đã có thể nâng khả năng gắn từ 400-700 kb (Paterson 1996).

Công nghệ cloning các megabase DNA vẫn tiếp tục được cải tiến. Hầu hết các tiến bộ gần đây đều tập trung phát triển kỹ thuật BAC. Khả năng gắn vào vectơ của BAC tuy không so được với YAC (những BAC đầu tiên gắn được 150 kb, và gần đây gắn được 350 kb), nhưng hệ thống BAC cho phép kỹ thuật cloning dễ thực hiện hơn, làm cho nó ngày càng trở nên phổ biến hơn.

Số lượng các clone cần thiết để xác suất tìm ra một clone mục tiêu (target clone) - được ước tính theo công thức như sau:

$$n = \ln(1 - \text{xác suất được yêu cầu}) / \ln(1 - \text{kích thước DNA được gắn vào} / S)$$

S = kích thước genome đơn bội thể

Ở genome người, thư mục YAC có 33.000 clone, với kích thước DNA gắn vào trung bình là 900 kb, nó có thể fingerprint bằng nhiều kỹ thuật để phối hợp các đoạn phân tử DNA chồng lấp nhau, những đoạn này đều đã lướt qua trên mỗi nhiễm thể (Cohen và ctv.

1993). Trong *Arabidopsis*, có 374 YAC clone đã được sắp xếp trong bốn contig, chạy trên nhiễm thể số 4 (Schmidt và ctv. 1995).

7.3.7. Tăng cường số marker bổ sung những khiếm khuyết

Việc xây dựng bản đồ contig hoàn chỉnh còn nhiều hạn chế, đối với một nhóm loài trong phân loại thực vật, ít nhất là trong một tương lai gần, nó phải được ứng dụng một cách hữu hiệu cho các nhà chọn giống cây trồng. Hai cây được tập trung cải tiến và có nhiều báo cáo là *Arabidopsis* và lúa. Việc lập bản đồ vật lý trên qui mô lớn của hai loài này đang được thực hiện, với nhiều nỗ lực cải tiến không ngừng. Bởi vì bộ genome chúng có kích thước nhỏ hơn so với các loài khác, để xây dựng thông số kỹ thuật có tính tiêu chuẩn hoá.

Để thu nhận được số lượng lớn các DNA marker cần cho vị trí đặc biệt của nhiễm thể nào đó, người ta thường bổ sung thêm phương pháp thanh lọc trên cơ sở PCR đối với những DNA marker bổ sung tại vùng mục tiêu của nhiễm thể đó. Quần thể dòng NIL (gần đẳng gen) hay dòng DNA pools đã được khuyến cáo để phát hiện tính đa hình, trong chẩn đoán các đoạn phân tử DNA liên kết chặt với gen mục tiêu. Người ta mất nhiều năm mới có thể sản xuất được các dòng NIL như vậy, cho nên DNA pools tổng hợp có thể được xem xét là phương pháp "tăng cường số marker bổ sung cho vùng mục tiêu". Kỹ thuật sản xuất này có protocol riêng, được công bố (Paterson 1996).

Một số lượng lớn các DNA marker trong vùng mục tiêu thật sự rất có ích, trừ khi một marker nào đó đủ để xác định liên kết với gen mục tiêu trên nhiễm thể - mà điều này rất hiếm. Gần đây, kỹ thuật gộp chung DNA thứ cấp (secondary DNA pooling) được mô tả như một giải pháp hữu hiệu cho kỹ thuật mapping các vùng đặc biệt trên nhiễm thể (Churchill và ctv. 1993). Bằng cách kết hợp "phương pháp đa dạng nguồn marker" với "phương pháp mapping có hiệu quả cao", tiến trình chromosome walking có thể tránh được, không cần thiết phải mất nhiều thời gian, và người ta có thể xác định marker liên kết chặt với một gen mục tiêu nào đó, gen này có thể định vị (landing) trên một megabase DNA clone (Tanksley và ctv. 1995).

7.3.8. Phương pháp phân lập gen bằng kỹ thuật Map-based cloning

Tập hợp contig, hoặc chromosome walking/landing đều là những kỹ thuật hỗ trợ cho việc tạo dòng vô tính trên cơ sở bản đồ, và sự xác định gen trên cơ sở thông tin của bản đồ di truyền. Kỹ thuật có tên là "map-based cloning" là một phương tiện để phân lập gen dựa trên sự phân ly của những alen đột biến trong thiên nhiên, hay đột biến nhân tạo. Kỹ thuật này cho phép chúng ta có thể phân lập gen trong điều kiện có ít thông tin. Kỹ thuật map-based cloning là phương tiện phân lập gen có hiệu quả, nhưng nó trở nên phức tạp khi genome quá lớn, có nhiều đoạn phân tử DNA có tính lặp lại, và đa bội thể. Vì vậy, người ta đã cố gắng nhân bản vô tính các gen quan trọng của thực vật bằng nhiều phương tiện như

phát sinh đột biến dính bên trong (insertional mutagenesis) (Paterson 1996).

Kỹ thuật map-based cloning có những yêu cầu cơ bản như sau :

- Phác họa gen mục tiêu thành những quãng ngắn trên nhiễm thể, luôn có hai marker kề cận gen, và nó được quét (spanning) bởi một dòng megabase DNA, hoặc bởi một contig của nhiều megabase DNA.
- Phương tiện để xác định việc chuyển mã trong megabase DNA.
- Hệ thống chuyển nạp có hiệu quả trong sự kiện du nhập DNA ngoại sinh vào loài thực vật đang nghiên cứu, cho phép chúng ta xác định gen mục tiêu bằng hiện tượng "bổ sung thể đột biến" (mutant complementation).

Việc phác họa gen mục tiêu thành những quãng ngắn trên nhiễm thể là một khối lượng to lớn. Thanh lọc một số lượng lớn các marker để phát hiện cái nào có liên kết với gen mục tiêu, với một lượng lớn các kiểu gen tái tổ hợp, sẽ giúp cho chúng ta một giải pháp tối ưu về bản đồ, nó hướng dẫn mọi nghiên cứu về megabase DNA có chứa gen mục tiêu này. Khi bản đồ contig đã sẵn sàng, thì việc nghiên cứu các dòng megabase DNA sẽ được thực hiện có hiệu quả hơn. Ngược lại, người ta phải sử dụng một local contig nào đó để chạy chromosome walking.

Cho đến nay, việc phân lập thành công những sequence đã được chuyển mã từ vùng mục tiêu của DNA, bao gồm sự kiện kết hợp nhiều phương pháp, chưa có trường hợp nào chỉ sử dụng một phương pháp mà thành công (Gardiner và Mural 1995). Trong trường hợp genome có kích thước lớn như cà chua, người ta kết hợp phương pháp cổ điển : thanh lọc trực tiếp thư mục cDNA, với phương pháp megabase DNA, kết quả rất thành công (Martin và ctv. 1994). Gần đây, người ta áp dụng PCR để thực hiện phương pháp megabase DNA có tính chất từng cá thể để kết dính một cách có chọn lọc với cDNA, sau đó cho khuếch đại nhờ phản ứng PCR (Parimoo và ctv. 1991, Lovett 1995).

7.4. ÁP DỤNG TRÊN CÂY LÚA (CASE STUDY)

Ngày nay, việc lập bản đồ gen cây lúa trở nên dễ dàng, nhanh, và đơn giản, nhờ phương tiện DNA marker. Các bước cơ bản trong chương trình thành lập bản đồ gen diễn ra như sau :

- Chọn một tính trạng nào đó.
- Chọn bố mẹ có sự tương phản rất rõ về tính trạng này [thí dụ mức độ nhiễm và kháng bệnh].
- Chuẩn bị các quần thể để làm bản đồ.
- Đo lường tính trạng nghiên cứu trên quần thể mapping.

- Phân tích kiểu gen về marker có tính đa hình trong quần thể mapping.
- Phân tích liên kết gen để xác định tính chất liên kết có chặt chẽ hay không.

7.4.1. Chọn một tính trạng nghiên cứu

Tính trạng được chọn để nghiên cứu trong bản đồ gen phải có giá trị quan trọng về nông học, bởi vì việc thực hiện rất tốn kém và nguồn vật liệu hiện còn rất hạn chế. Tính trạng này có thể được ứng dụng trong kỹ thuật được gọi là "chọn lọc nhờ marker" sau này. Người ta còn xem xét các tiêu chuẩn khác như : mức độ hiểu biết về nó ra sao, khả năng tạo ra vật liệu, và vị trí của gen trên nhiễm thể. Nếu vị trí của gen trên nhiễm thể đã được biết thì việc xây dựng bản đồ sẽ trở nên đơn giản hơn.

7.4.2. Chuẩn bị quần thể làm bản đồ

Sau khi chọn lựa kỹ lưỡng bố mẹ, chúng ta tiến hành lai và hình thành các quần thể sau đây:

- Quần thể đơn bội kép [DH] : double haploid
- Quần thể cận giao tái tổ hợp [RI] : recombinant inbred
- Các dòng gần như đẳng gen [NIL] : nearly isogenic lines
- Quần thể F₂ tiêu chuẩn
- Quần thể F₂ và F₃
- Quần thể hồi giao [BC₁] : backcross

Các quần thể khác như lai ba thử nghiệm (test cross hay triple cross) cũng có thể được sử dụng. Việc sử dụng quần thể nào để lập bản đồ gen còn tùy thuộc vào khả năng của vật liệu và kinh nghiệm nhà nghiên cứu.

7.4.3. Chọn lựa bố mẹ

Tiêu chuẩn quan trọng nhất trong chọn lựa bố mẹ là phải có sự tương phản rất lớn về tính trạng nghiên cứu có trong bố mẹ và sự di truyền của tính trạng này ở bố và mẹ. Nếu sự phân ly về kiểu hình thực sự rõ ràng, thì việc lập bản đồ gen mới có tính khả thi.

Xét về mức độ đa hình giữa bố mẹ, tính chất này càng lớn, kết quả càng tốt. Tính đa hình giữa các giống indica/japonica thường cao (khoảng 70%) nhưng nó có trở ngại rất lớn là tính bất thụ. Quần thể này nên có độ lệch rất ý nghĩa giữa bố mẹ, mà bố hoặc mẹ này có tính thích nghi với môi trường trong đó quần thể đang phát triển.

Qui mô quần thể phải đủ lớn cho phép chúng ta ước đoán khá chính xác giá trị liên kết gen. Tuy nhiên quần thể lớn cỡ nào còn tùy thuộc vào giá chi phí phải tốn kém. Bởi vì chi phí để thực hiện rất đắt, nên trước đây, người ta sử dụng quần thể có qui mô nhỏ để lập

bản đồ gen. Đối với các tính trạng đơn gen, qui mô quần thể theo tiêu chuẩn GML tại IRRI là 100-150 cá thể / quần thể.

7.4.4. Đo lường kiểu hình

Nhìn chung, tính trạng đơn gen rất dễ quan sát và đo đếm, nhưng không phải luôn luôn trong mọi trường hợp. Kiểu hình là kết quả của ảnh hưởng giữa kiểu gen và môi trường. Do đó, điều rất quan trọng là phải làm sao đo đếm một cách chính xác kiểu hình. Người ta sử dụng một quần thể trong đó cho phép kiểu hình được lập lại, điều này có lợi là làm tăng độ chính xác khi đo đếm, đặc biệt đối với những tính trạng mẫn cảm đối với sự thay đổi do môi trường. Các dòng DH hoặc RI, rất hữu dụng trong vài trường hợp. Nếu chỉ có F₂, chúng ta hãy cố gắng lấy DNA từ những cá thể F₂ và ghi nhận các số liệu về kiểu hình của F₂. Bất kỳ trường hợp nào, việc phân tích kiểu hình phải là công việc được đầu tư nhiều nhất. Điều quan trọng hơn là phải xác định cho được chất lượng của việc lập bản đồ gen.

7.4.5. Phân tích kiểu gen

Phân tích kiểu gen là nhiệm vụ trung tâm của công tác lập bản đồ gen. Trước hết khảo sát bố mẹ, sau đó sử dụng các marker đa hình để khảo sát quần thể phân ly. Người ta có thể sử dụng cả hai phương pháp phân tích Southern và PCR. Vì phương pháp phân tích kiểu gen rất đắt tiền, nên có nhiều tác giả đã cải tiến, rút gọn qui trình. Cách thông dụng nhất là lập bản đồ gen tại các vị trí của nhiễm thể đã biết trước, hoặc sử dụng các dòng NIL, hoặc sử dụng kỹ thuật phân tích quần thể phân ly trồng dồn (bulk segregant analysis).

7.4.6. Phân tích liên kết gen

Phân tích liên kết gen giữa hai marker là công việc rất đơn giản, có thể thực hiện bằng phương pháp thủ công. Tư liệu hóa các liên kết gen giữa những marker cần được sự trợ giúp của *puter*. Người ta đã sử dụng chương trình MAPMARKER để lập bản đồ các marker.

7.4.7. Yêu cầu rút ngắn qui trình đánh dấu một gen đơn

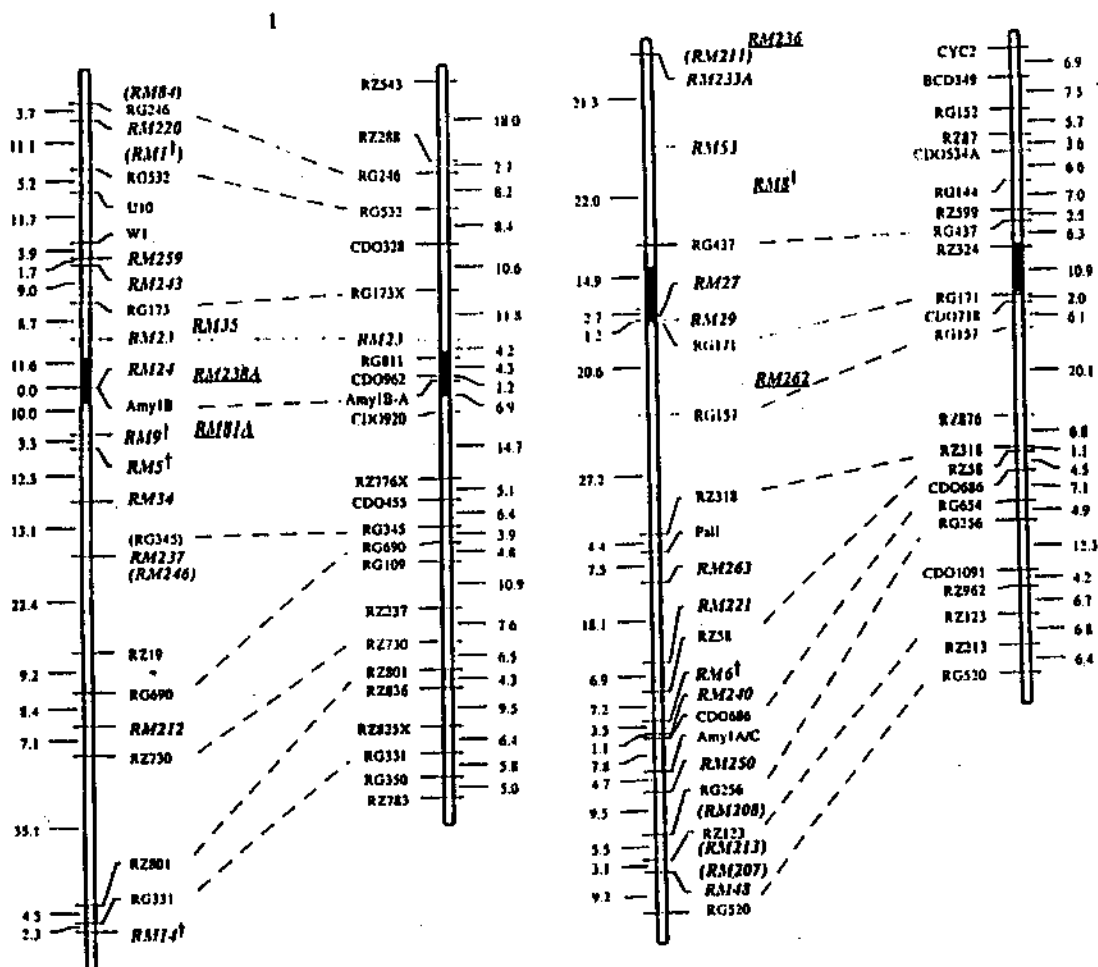
Mặc dù người ta đã tin tưởng rằng RFLP marker có thể phủ kín genome cây lúa, nhưng cho đến nay công việc phân tích marker đối với mỗi tổ hợp lai cụ thể, vẫn còn là công việc rất lớn. Do đó người ta vẫn phải tiếp tục tìm kiếm các qui trình đơn giản hơn và thuận lợi hơn. Thí dụ, thay vì sử dụng RFLP, người ta có thể sử dụng các marker STS, hoặc microsatellite, với qui trình đơn giản hơn rất nhiều (hình 7-4)

Phương pháp lập bản đồ gen ở vị trí của nhiễm thể được biết trước :

Thí dụ lập bản đồ gen *Xa-4* : người ta tạo ra một quần thể từ cặp lai Ma Hae x IR36. Đánh giá kiểu hình ở F₂ và F₃ nhằm đảm bảo mức độ ghi nhận tính kháng có chính xác hay không.

Trước khi bắt đầu mapping, người ta được biết gen này định vị ở nhiễm thể số 11. Do đó, người ta chỉ có chọn lọc các marker trên nhiễm thể 11 của bố mẹ nhằm xác định các marker có tính chất đa hình. Sau khi phân tích liên kết gen, người ta thấy rằng *Xa-4* nằm giữa hai RFLP marker, đó là RZ536 và RG303 (Hình 7-5). Tương tự, gen *xa-5* trên nhiễm thể 5.

Phương pháp này rất tiện lợi, nó làm giảm số lượng RFLP marker hơn 90%.



Hình 7.4 : Bản đồ di truyền cây lúa : so sánh giữa RFLP (phải) và microsatellite marker (trái) ở nhiễm thể 1 và 2, trên cơ sở quần thể đơn bội kép (DH) của IR64/Azucena. Microsatellite marker có LOD > 2 thống nhất với khung của RFLP, marker nào đó có LOD < 2 [trong dấu ngoặc đơn] định vị ở các vị trí có xác suất cao nhất dọc theo nhiễm thể (Chen và ctv. 1997)

Lập bản đồ gen với các dòng NIL :

Dòng đẳng gen hay dòng gần như đẳng gen (NIL) là gì ?

NIL được tạo ra do hồi giao giữa con lai với một trong hai bố mẹ của nó. Do đó nó

hình thành một cặp của các dòng NIL, và nó chỉ khác nhau ở một vài đoạn mã di truyền của bố mẹ cho gen. Thí dụ giống IR 24 và các dòng NIL của nó mang gen kháng BLB (bệnh bạc lá lúa).

Muốn lập bản đồ gen với các dòng NIL, chúng ta phải thực hiện các bước như sau :

1. Chọn các gen để lập bản đồ (thí dụ gen *xa-5* và *BlB*).
2. Chọn bố mẹ (một cặp của các dòng NIL).
3. Lai để có quần thể F2 phân ly.
4. Phân lập DNA của mỗi cá thể F2 và bố mẹ của chúng.
5. Chủng bệnh trên từng cá thể ở F2 và bố mẹ để ghi nhận phản ứng kháng.
6. Chọn các marker trên cơ sở bản đồ RFLP, sao cho mỗi marker riêng biệt nhau với khoảng cách 20cM.
7. Khảo sát tính chất đa hình với các marker đã được chọn lọc trên hai bố mẹ (Hình 4-2)
8. Sử dụng 6 restriction enzyme để tạo cơ hội có được thể đa hình.
9. Cho điểm tất cả cá thể F2 bằng RFLP.
10. Thu thập số liệu.

Hoàn thành việc phân tích liên kết gen bằng computer, chương trình MAPMARKER.

Không cần phải phân tích Southern blot trên những marker có tính chất đa hình mà không du nhập vào được. Thí dụ trong trường hợp vẽ bản đồ gen *xa-5*, tổng số có 4 vùng được phát hiện, và 53 cá thể F2 được phân tích. Chỉ có một vùng trên nhiễm thể số 5 có liên kết với gen *xa-5* (Hình 7-5).

Với phương pháp này, bất cứ một thể đa hình nào có tương quan với mức độ du nhập gen từ thể cho [donor] sang thể nhận [recipient], đều được tìm thấy. Khối lượng công việc khảo sát các con lai, giảm đi một cách có ý nghĩa. Nhưng bù lại chúng ta phải mất nhiều thời gian để tạo ra các dòng NIL.

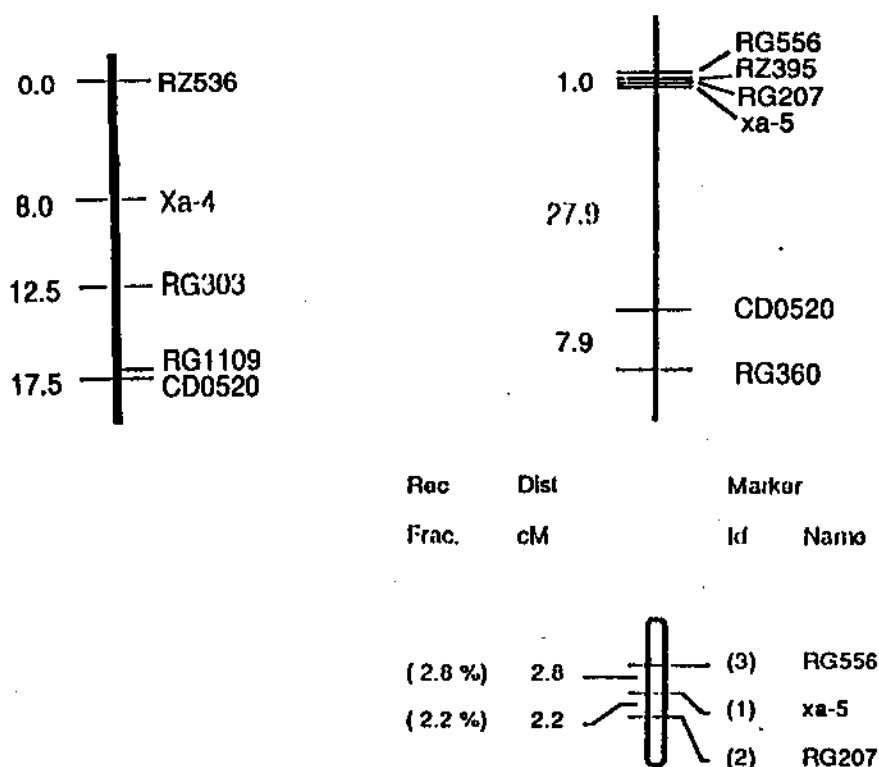
Phân tích các dòng phân ly trồng đồn [BSA] :

BSA được viết tắt từ chữ " bulk segregant analysis". Tiến trình có thể xem như khảo sát các dòng NIL giả [pseudo-NILs] trong quần thể F2 :

1. Chọn các gen để lập bản đồ.
2. Chọn bố mẹ.
3. Lai để có quần thể F2.

4. Phân lập DNA từ những cá thể của F2 và bố mẹ của chúng.
 5. Hoàn thành phân tích kiểu hình.
 6. Chọn các marker trên cơ sở bản đồ RFLP sao cho mỗi marker tách biệt nhau ở khoảng cách chừng 20cM.
 7. Trồng F3 theo nhóm dòng để xác định tính chất đồng hợp tử của các cá thể F2.
 8. Tạo ra hai quần thể trồng dần từ hai nhóm đồng hợp tử.
 9. Hoàn thành việc khảo sát tính chất đa hình, với những marker được chọn lọc trên hai bố mẹ, tiến hành đồng thời với quần thể trồng dần.
 10. Cho điểm tất cả các cá thể F2 bằng polymorphic marker.
 11. Hoàn thành phân tích liên kết gen bằng computer, chương trình MAPMARKER.
- Hiện nay, người ta sử dụng RAPD trong BSA, nó sẽ cho kết quả nhanh hơn rất nhiều

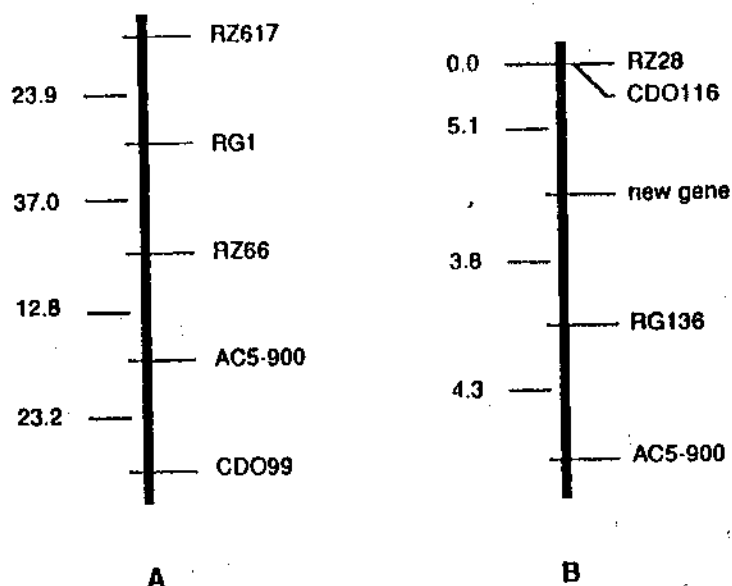
lần.



Hình 7.5 : Bản đồ liên kết gen không hoàn toàn của nhiễm thể 11 của cây lúa, đối với gen Xa-4 và nhiễm thể 5 đối với gen xa-5, trên cơ sở phân tích ALP, dùng MAPMARKER để vẽ bản đồ.

7.4.8. Lập bản đồ RAPD và RFLP đối với tính kháng bệnh bạc lá lúa

Bệnh bạc lá lúa do vi khuẩn *Xanthomonas oryzae* pv. *oryzae* (Xoo) - là một trong những bệnh hại nghiêm trọng nhất trên cây lúa. Pathogen thay đổi độc tính của nó tùy theo giống lúa. Người ta đã phân lập được sáu nòi [races] vi khuẩn này tại Philippines. Mỗi nòi có độc tính riêng trên các giống khác nhau với những gen kháng khác nhau, theo khái niệm "gen đối gen" trong mối quan hệ giữa ký chủ và ký sinh. Cho đến nay, có hơn 18 gen kháng *Xa* đã được phân lập bằng phương pháp phân tích di truyền cổ điển. Một gen mới, có tính lặn, kháng với nòi số 6 của *Xoo* đã được tìm thấy tại Philippines.



Hình 7.6 : Bản đồ RFLP trên nhiễm thể 8 của cây lúa, khoảng cách tính bằng cM. [A] bản đồ được vẽ trên cơ sở quần thể vật liệu của IRRI. Marker AC5-900 nằm giữa hai RFLP marker là RZ66 và CDO99. [B] bản đồ được phát triển trên cơ sở 132 cây F2 của cặp lai IRBB13/IR24, cho thấy vị trí của gen mới liên kết chặt chẽ với marker.

Vật liệu : Dòng gần như đẳng gen đối với gen mới này, dòng cho gen kháng (hạt dài, mã số Acc.35023), và dòng nhiễm IR24 làm giống tái tục, được dùng làm vật liệu nghiên cứu. Dòng NIL được tạo ra từ cặp lai giữa IR24 / Giống hạt dài (Long Grain), qua năm lần hồi giao với giống tái tục IR24, tiếp theo đó là 5 lần tự giao phối. Quần thể phân ly F2 của 132 cây được tạo ra bằng cách lai giữa NIL và IR24.

Phản ứng đối với bệnh bạc lá: Bố mẹ và con lai F2 được trồng trong nương mạ, xây bằng khung bê tông. Vào giai đoạn đẻ chồi tối đa, cây lúa được chủng PX099 [nòi số 6] của *Xoo*, theo phương pháp cấy lá. 14 ngày sau đó, quan sát vết bệnh và cho điểm, trên cơ sở phần trăm diện tích vết bệnh so với diện tích lá.

Ly trích DNA, phản ứng phân cắt DNA, điện di và phân tích Southern :

DNA của cây lúa được chuẩn bị từ mô lá tươi, đông khô, theo Dellaporta và ctv. (1983). Genomic DNA của IR24, giống hạt dài, quần thể trồng đồn mang tính kháng, quần thể trồng đồn mang tính nhiễm, được tiêu hóa bằng các restriction enzyme như *Bam*HI, *Bgl*II, *Dra*I, *Eco*RI, *Eco*RV, *Hind*III, *Pst*I, *Sal*I, *Sac*I, *Xba*I và *Xho*I. Thể trồng đồn mang tính kháng và thể trồng đồn mang tính nhiễm được trích DNA, từ 10 cây kháng và 10 cây nhiễm trong F2. Đối với các vật liệu lọc con lai F2 (progeny filters), người ta chỉ sử dụng các enzyme có phản ứng dương tính. Điện di và phân tích Southern được thực hiện theo nguyên tắc của Sambrook và ctv. (1989).

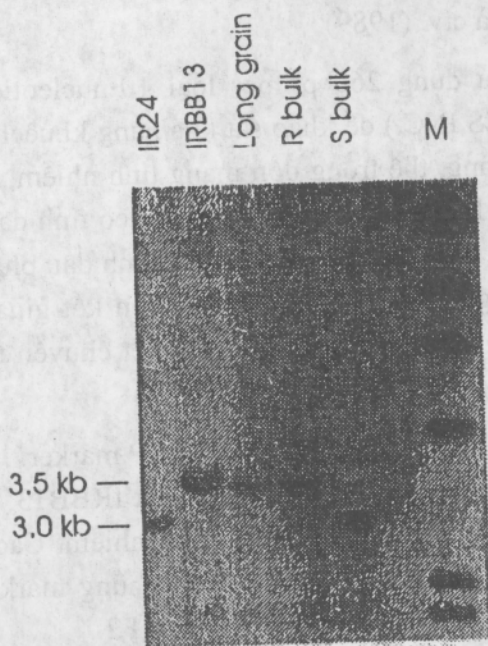
Phân tích RAPD : sử dụng 260 primer loại 10-nucleotide (Operon 10-mer Kits, OPERON TECHNOLOGIES, INC.) để khảo sát khả năng khuếch đại các vệt trong bố mẹ, thể trồng đồn mang tính kháng, thể trồng đồn mang tính nhiễm, theo phương pháp RAPD của William và ctv. (1990), Martin và ctv. (1991). Vệt có tính đa hình và dương tính được phân lập, và được làm thuần trên tấm gel, sau đó cho đánh dấu phóng xạ bằng phương pháp primer ngẫu nhiên (Feinberg và Vogelstein 1983). Liên kết giữa RAPD marker và locus của gen mới này được phân tích bằng cách sử dụng vệt chuyên tính như chất thăm dò để khảo sát các thể lọc của 132 cây F2.

Phân tích liên kết giữa locus của gen mới và RFLP marker: 10 RFLP marker có quan hệ rất gần với RAPD marker được chọn lọc để khảo sát IRBB13, giống hạt dài, IR24, thể trồng đồn mang tính kháng, thể trồng đồn mang tính nhiễm. Các RFLP marker cho thấy tính chất đa hình giữa dòng kháng và dòng nhiễm. Những marker này được sử dụng để phân tích liên kết của nó với gen mới trong 132 cây của F2

Phân tích bằng computer : người ta tính các giá trị liên kết theo phần mềm MAPMARKER của Lander và ctv. (1987). Khoảng cách giữa các marker được biểu thị bằng centi Morgan, theo hàm Kosambi (1944). Liên kết thực sự có ý nghĩa, nếu lod (logarithm of odds) ở khoảng ≥ 3.0 . Người ta tính toán khoảng cách di truyền này bằng cách phân tích nhiều điểm (multipoint analysis).

Lập bản đồ RAPD của locus gen mới: Người ta sử dụng phân tích RAPD như bước khởi đầu cho việc lập bản đồ của gen mới. Theo kết quả điều tra 260 random primer, người ta phân lập được một primer AC5 có tác dụng khuếch đại vệt đa hình giữa những cây kháng [bố mẹ kháng, thể cho, thể trồng đồn có tính kháng] và những cây nhiễm [bố mẹ nhiễm, thể trồng đồn có tính nhiễm]. Một vệt có qui mô 0.9 kb được khuếch đại từ cây nhiễm IR24 và thể trồng đồn có tính nhiễm. Dấu vết của vệt này cũng được tìm thấy trong thể trồng đồn có tính kháng (Hình 7.7 và 7.8). Lúc DNA của 10 cây kháng và 10 cây nhiễm trong F2 được khuếch đại bằng primer AC5, thì vệt 0.9 kb được tìm thấy ở tất cả các cây nhiễm, nhưng chỉ được tìm thấy trên hai cây kháng. Để khẳng định liên kết này [giữa locus của gen mới với RAPD - AC5-900], người ta đã phân lập vệt chuyên tính trên tấm gel, sử dụng nó như một thể thăm dò có tính chất lai (hybridization probe) đối với các

genomic DNA đã bị phân giải bằng 12 restriction enzyme. Các đoạn DNA có tính chất đa hình đã được phát hiện có trong giống nhiễm IR24 và cây kháng (IRBB13, giống hạt dài, thể trồng đồn có tính kháng). Thể trồng đồn có tính nhiễm thể hiện cả hai loại vệt, bởi vì nó có chứa một vài dị hợp tử (Hình 7-7). Sử dụng chất thăm dò AC5-900 để khảo sát các thể lọc có chứa DNA được ly trích từ 132 cây F2, người ta đã tìm ra khoảng cách giữa AC5-900 locus và locus gen mới là 5.3 cM.



Hình 7.7 : Phim chụp cho thấy hiệu quả lai giữa RG136 và DNA được phân cắt bởi *DraI* để phát hiện gen kháng xa-13 trên cây lúa.

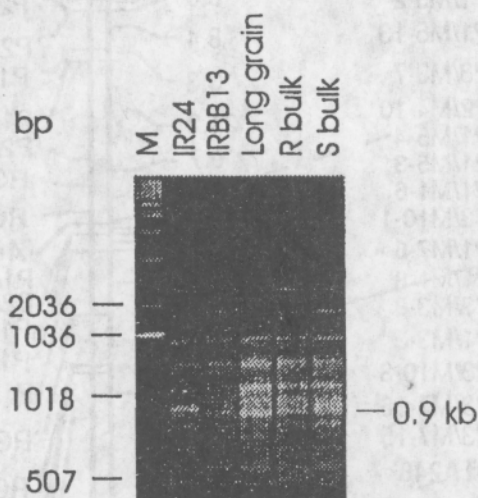
Vị trí của RAPD trên bản đồ RFLP :

Chất thăm dò AC5-900 được sử dụng để khảo sát các filter, trong đó có *ScaI* lấy từ quần thể mapping của IRRI, sau khi khảo sát tính đa hình DNA của bố mẹ. Người ta ghi nhận các kiểu gen của những đoạn mã này. Liên kết giữa RAPD marker và RFLP marker trong quần thể mapping cũng được tính toán. Kết quả cho thấy RAPD marker được định vị giữa các RFLP marker là RZ66 và CD099, trong nhiễm thể số 8 (Hình 7-6).

Phân tích liên kết gen giữa RFLP marker và gen mới :

10 RFLP marker phản ứng với AC5-900 locus đã được chọn lọc để lai các DNA blot của giống IRBB13, giống hạt dài, IR24, của quần thể trồng đồn mang tính kháng, và của quần thể trồng đồn mang tính nhiễm. Người ta dùng 12 restriction enzyme, để xem thể đa hình của những dòng nhiễm và dòng kháng, với sự phát hiện của RZ28 và CD0116, trong điều kiện DNA được phân cắt bởi *BglII*, *EcoRI*, *SacI*, *XbaI*, *XhoI*, và *DraI* (Hình 7-8). Bả marker còn lại RZ926, CD0565, RZ66, RZ549, RZ572, CD099, và RZ997 cho thấy tín

chất đơn hình. Người ta sử dụng RFLP marker để khảo sát các filter của 132 cây F2. Người ta không tìm thấy các con lai tái tổ hợp [recombinant] giữa RZ28 và CD0116. Khoảng cách di truyền giữa hai marker này và gen mới là 5.1 cM. Trong khi RG136 nằm trên một vị trí khác của gen mới, khoảng cách của nó với gen mới là 3.8 cM (Hình 7-6).

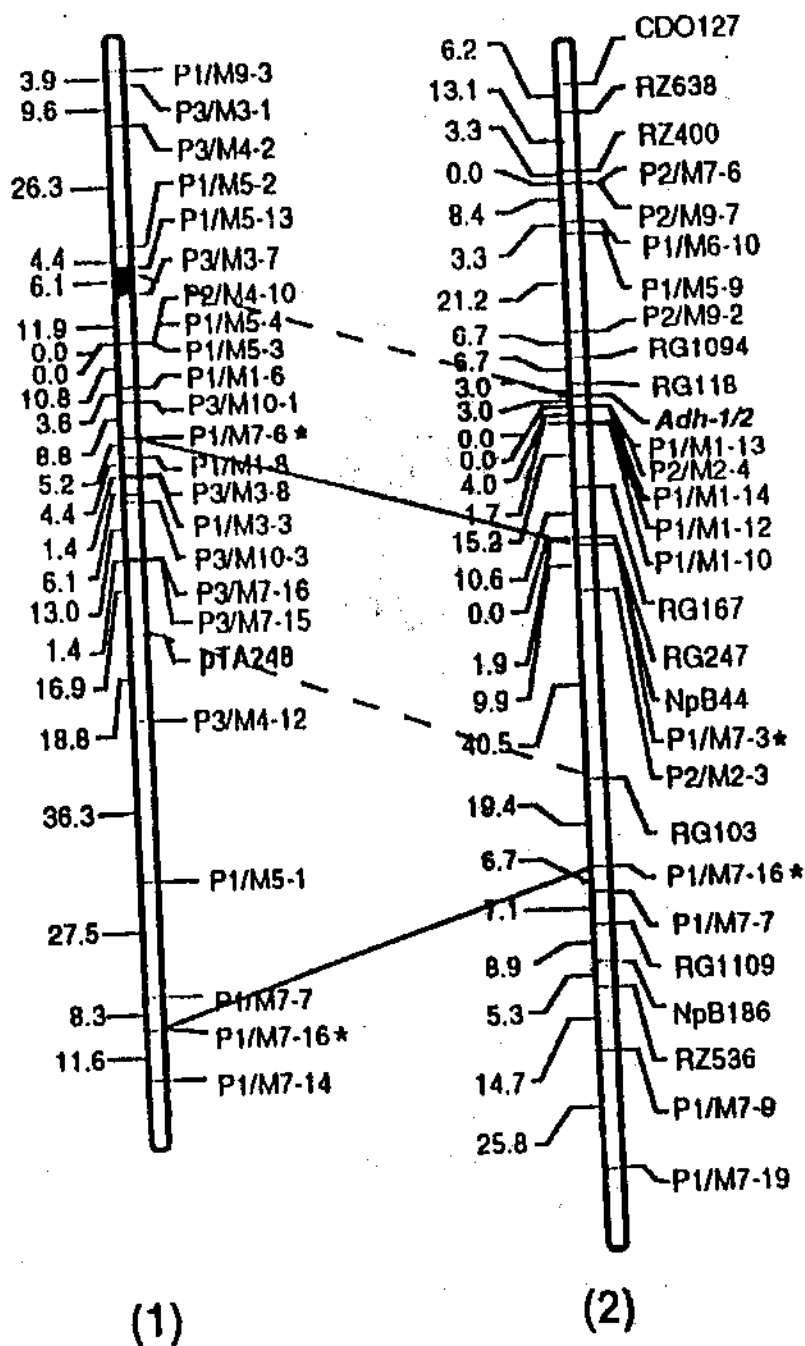


Hình 7.8 : Điện di sản phẩm PCR, sử dụng primer ngẫu nhiên AC5 cho thấy mức độ đa hình PCR. Nó cũng có thể giúp chúng ta phát hiện gen xa-13 trên quần thể trồng đồn mang tính kháng và quần thể trồng đồn mang tính nhiễm.

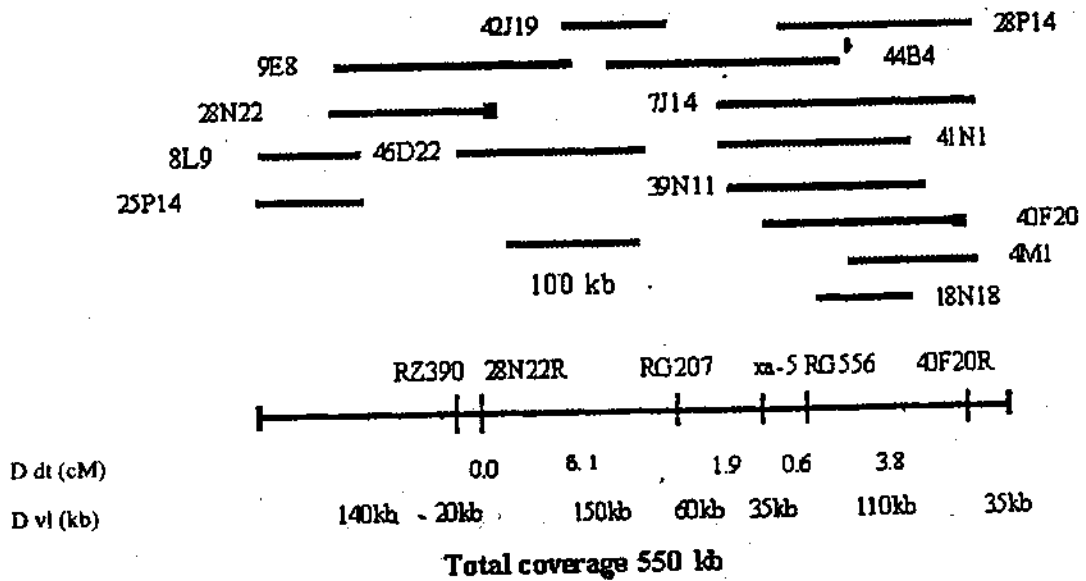
Đối với các tính trạng QTL, như phân tích tính chống chịu ngập hoàn toàn của cây lúa, người ta đã sử dụng AFLP để phát hiện gen chống chịu (hình 7-9)

7.4.9. Bản đồ vật lý cây lúa

Genome cây lúa được xem như nhỏ nhất trong lớp thực vật một lá mầm. Khoảng cách vật lý : khoảng cách di truyền (DNA : cM) được ước lượng là 250-300 kb : cM (Arumuganathan và ctv. 1991, Kurata và ctv. 1994). Thư mục YAC (Umehara và ctv. 1995) và BAC (Wang và ctv. 1995, Zhang và ctv. 1995) đã được liên tục nghiên cứu và ứng dụng. Bên cạnh đó cosmid cũng đã được công bố (Song 1995). Bản đồ vật lý trên cơ sở YAC contig (Kurata và ctv. 1994, Sasaki và ctv. 1994) đã được xây dựng và được bổ sung liên tục do Tiến sĩ Sasaki lãnh đạo, trong chương trình Rice Genome tại Nhật là một thành tựu khá nổi bật trong những năm gần đây (hình 7-1: bản đồ vật lý trên cơ sở YAC và hình 7-10: bản đồ vật lý trên cơ sở BAC). Đối với các tính trạng di truyền số lượng, như phân tích gen chống chịu ngập hoàn toàn của cây lúa, người ta đã sử dụng AFLP để mapping các gen này (hình 7-9).



Hình 7.9 : Bản đồ di truyền ở nhiễm thể 11 của cây lúa, trên cơ sở : (1) quần thể cận giao tái tổ hợp [RIL] giữa IR74/FRI3A, và (2) quần thể đơn bội kép [DH] của tổ hợp lai này. Gạch nối chỉ vị trí của gen *Adh1/2* và các gen QTL, kiểm soát tính chống chịu ngập trên cây lúa (Nandi và ctv. 1995).



Ddt = khoảng cách di truyền, Dvl = khoảng cách vật lý.

Hình 7.10 : Bản đồ vật lý sử dụng BAC clone phủ trên gen *xa-5* ở một nhánh gần của nhiễm thể 5. Khoảng cách di truyền được tính theo phân tích Southern. Marker 28N22R và 40F20R được tạo ra nhờ Tail PCR từ đầu bên phải của BAC clone, được biểu thị bằng ô vuông trên hình. Những clone chồng lấp lên nhau được biểu thị bằng những gạch đậm (Yang 1997)

TÀI LIỆU THAM KHẢO

1. Allard RW. 1959. *Formulas and tables to facilitate the calculation of recombination values*. Heridity. 24: 235-279
2. Anderson J, Y Ogihara, ME Sorellis, SD Tanksley. 1992. *Development of a chromosome arm map for wheat based on RFLP markers*. Theo. Appl. Genet 83: 1035-1043
3. Arondel V, B Lemieux, I Hwang, S Gibson, HM Goodman, CR Someville. 1992. *Map-based cloning of a gene controlling Omega-3 fatty acid denaturation in Arabidopsis*. Science 258: 1353-1354
4. Arumuganathan K, ED Earle. 1991. *Nuclear DNA content of some important plant species*. Plant Mol. Biol. Rep. 9:208-218
5. Bonierbale MW, RL Plaisted, SD Tanksley. 1988. *RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato*. Genetics 120: 1095-1103
6. Brolin P, MW Ganai, SD Tanksley. 1992. *Telomic array display high levels of heritable polymorphism among closely-related plant varieties*. Proc. Natl. Acad. Sci. USA 89: 1354-1357
7. Burke DT, GF Carle, MV Olson. 1987. *Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors*. Science 236:806-811
8. Buu BC, NT Lang. 1997. *Rice germplasm evaluation assisted by DNA markers in Mekong Delta, Vietnam*. P. 267. Poster presentation abstract. General Meeting of The Int. Prog. On Rice Biotechnol.. 15-19 Sept. 1997, Malacca, Malaysia.
9. Carlson JE, LK Tulsieram, JC Glaubitz, VWK Luk, C Kauffelt, R Rutledge. 1991. *Segregation of RAPD markers in F1 progeny of Conifer*. Theo. Appl. Genet. 83: 194-200
10. Chang C, JL Bowman, AW Dejhon, ES Lander, EM Meyerovitch. 1988. *Restriction fragment length polymorphism linkage map of Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA 85: 6856-6860
11. Chen X, S Temnykh, Y Xu, YG Cho, SR McCouch. 1997. *Development of a microsatellite framework map providing genome-wide coverage in rice (Oryza sativa L.)*. Theor. Appl. Genet. 95:553-567
12. Cheung WY, MD Gale. 1990. *The identification of high molecular weight DNA from wheat, barley, and rye for analysis by pulsed field gel electrophoresis*. Plant Mol. Biol. 14:881-888

13. Churchill GA, JJ Giovannoni, SD Tanksley. 1993. *Pooled-sampling makes high resolution mapping practical with DNA markers*. Proc. Natl. Acad. Sci. USA 90:16-20
14. Cohen D, I Chumakov, J Weissenbach. 1993. *A first generation physical map of the human genome*. Nature 366:968-701
15. Cowen NM. 1988. *The use of replicated progenies in marker-based mapping of QTL's*. Theor. Appl. Genet. 75: 857-865
16. Cowen NM, CD Johnson, K Armstrong, M Miller, A Woosley, S Pescitelli, M Skokut, S Belmar, Petolono. 1992. *Mapping genes conditioning in vitro androgenesis in maize using RFLP analysis*. Theor. Appl. Genet. 84: 720-724
17. Dellaporta SL, J Wood, JB Hick. 1983. *A plant DNA mini preparation; Version II*. Plant Mol. Biol. Rep. 1:19-21
18. Devey ME, KD Jermstad, CG Tauer, DB Neale. 1991. *Inheritance of RFLP loci in a L lolly pine three-generation pedigree*. Theo. Appl. Genet. 83: 238-242
19. Edwards KJ, H Thompson, D Edwards, AD Saizieu, C Sparks, Thompson, AJ Greenland, M Evers, W Schuch. 1992. *Construction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents*. Plant. Mol. Biol. 19:299-308
20. Fedmann KA, MD Marks, ML Christianson, RSQuatrano. 1989. *A dwarf mutant of Arabidopsis generated by T-DNA insertion mutagenesis*. Science 243: 1351-1354
21. Ganai MW, ND Young, SD Tanksley. 1989. *Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato*. Mol. Gen. Genet. 215:395-400
22. Gardiner K, R Mural. 1995. *Getting the message: identifying transcribed sequences*. Trends Genet. 11:77-79
23. Gebhart C, E Ritter, A Barone, T Debener, B Walkemeler, U Schachtschabel, H Kaufmann, RD Thompson, MW Bonierbale, MW Ganai. 1991. *RFLP Maps of potato and their alignment with the homologous potato genome*. Theo. Appl. Genet. 83: 49-57
24. Giraudat J, BM Hauge, C Valon, J Smalle, F parcy, HM Goodman. 1992. *Isolation of the Arabidopsis AB13 gene by position cloning*. Plant Cell 4: 1251-1261
25. Graner A., A hoor, J Schondelmajor, H Siedler, K Pillen, G Fichbeck, G Wenzel, RG Herrmann. 1991. *Construction of an RFLP map of barley*. Theor. Appl. Genet. 83: 250-256

26. Gustafson JP, JE Dille. 1992. *Recombination linkage groups*. Proc. Natl.Acad. Sci. USA. 89: 8646-8650
27. Hauge BM, S Hansley, J Giraudat, HM Goodman. 1991. *Mapping the Arabidopsis genome*. Cambridge : Cambridge University press.
28. Honeycutt RJ, BWS S ral, M Mc Clelland, AG Athely.1992. *Analysis of large DNA from soybean (Glycine max L. Merr.) by pulsed field gel electrophoresis*. Plant J. 2:133-135
29. Hulbert SH, TE Richter, JD Extell, JL Bennetzen. 1990. *Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes*. Proc. Natl. Acad. Sci. USA. 97: 4251-4255
30. Hulbert SH., TW Ilott, EJ Legg, SE Lincol, ES Lander, RW Michelmore.1988. *Genetic analysis of the fungus Bremia lactucae using RFLP*. Genetics 120: 947-958
31. Kochert GT, Halward, WD Branch,CE Simpson. 1991. *RFLP variability in peanut I.Cultivars and wild species*. Theo. Appl. Genet. 81: 565-570
32. Kosambi DD. 1944. *The estimation of map distances from recombination values*. Ann. Eugen. 9:9-19
33. Kurata N, Y Nagamura, K Yamamoto, Y harushima, N Sue, J Wu, BA Antonio, A Shomura, T Shimizu, SY Lin, T Inoue, A Fukuda, T Shimano, Y Kuboki, T Toyama, Y Miyamoto, T Kirihara, K Hayasaka, A Miyao, L Monna, HS Zhong, Y Tamura, ZX Wang, T Momma, Y Umehara, M Yano, T Sasaki, Y Min e. 1994. *A 300 kilobase interval genetic map of rice including 883 expressed sequences*. Nature Genetics 8:365-372
34. Lang NT, PK Subudhi, SS Virmani, Ning Huang, DS Brar. 1997. *Development of PCR-based markers for thrmosensitive genetic male sterility tms-3(t) in rice*. P. 146. *Presentation abstract*. General Meeting of The Int. Prog. On Rice Biotechnol.. 15-19 Sept. 1997, Malacca, Malaysia.
35. Lander ES, P Green, J Abrahamson, A Barlow, MJ Daly, SE Lincol, L Newburg. 1987. *MAPMARKER: An interactive computer package for constructing primery genetic linkage maps of experimental and natural populations*. Genomics 1:174-181
36. Li Z, SRM Pinson, WD Park, AH Paterson, JW Stansel. 1996. *Mapping of quantitative trait loci and quantitative-trait-modifying factors affecting three grain yield components in rice (Oryza sativa L.)*. P. 624-630 in Proc. of the 3rd Int. Rice. Genet. Symp. Rice Genetics III. IRRI, Philippines.
37. Lovett M. 1995. *Fishing for compliments: finding genes by direct cDNA selection*. Trends Genet.

38. Martin GB, JGK Williams, SD Tanksley. 1991. *Rapid identification of markers linked to Pseudomonas resistance gene in tomato using random primers and near isogenic lines*. Proc. Natl. Acad. Sci. USA 88: 2336-2340
39. Martin GB, JGK William, SD Tanksley. 1991. *Rapid identification of markers linked to a Pseudomonas resistance gene in tomato by using random primers and near-isogenic lines*. Proc. Acad. Sci. USA 88:2336-2340
40. Martin GB, MW Ganai, SD Tanksley. 1992. *Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two distance resistance loci*. Mol. Gen. Genet. 223: 25-32
41. Martin GB, SH Brommonschenkel, J Chunwongse, A Frary, MW Ganai, R Spivey, T Wu, ED Earle, SD Tanksley. 1994. *Map-based cloning of a protein kinase gene conferring disease resistance in tomato*. Science 262:1432-1436
42. Michelmore RW, I Paran, RV Kesseli. 1991. *Identification of markers linked to disease resistance genes by bulked segregant analysis : a rapid method to detect markers in specific genomic regions using segregation populations*. Proc. Natl. Acad. Sci. USA. 88: 9828- 9832
43. Nam HG, J Giraudat, BD Boer, F Moonan, WDB Loos, BM Hauge, HM Goodman. 1989. *Restriction fragment length polymorphism linkage map of Arabidopsis thaliana*. Plant Cell 1: 699-705
44. Nandi S, PK Subudhi, D Senadhira, L Norvie, SS Mandi, N Huang. 1995. *Mapping QTLs with AFLP and selective genotyping for submergence tolerance in rice*. Chapter 3. DNA marker-facilitated QTL x Envirimental Analysis in Rice. Nov.20- 29Dec. 1995, IRRI, Philippines.
45. Parimoo S, SR Patam li, H Shukla, DD Chaplin, SM Weissman. 1991. *cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments*. Proc. Natl. Acad. Sci. USA 88:9623-9627
46. Paterson AH, ES Lander, JD Hewitt, S Peterson, SE Lincoln, SD Tanksley. 1988. *Resolution of quantitative traits into Mendelian factors by using a complete linkage map restriction fragment length polymorphisms*. Nature 335:721-726
47. Paterson AH. 1996. *Genome Mapping in Plants*. RG LANDES Co., Austin, USA
48. Paterson AH, and RA Wing. 1993. *Genome mapping in plant*. Biotechnology (4): 142-147
49. Paterson AH, SD Tanksley, ME Sorellis. 1992. *DNA markers in crop improvement*. In
50. Retter RS, JGK William, KA Fredman, Rafalski, SV Tingey, PA Scolnik. 1992. *Gl al and local genome mapping in Arabidopsis thaliana by using recombinant inbred*

51. Sasaki T, J Song, Y Koga-Ban, E Matsui, F Fang, H Higo, H Nagasaki, M Hori, M Miya, E Murayama, T Takiguchi, A Takasuga, T Niki, K Ishimaru, H Ikeda, Y Yamamoto, Y Mukai, I Ohta, N Miyadera, I Havukkala, Y Min e. 1994. *Toward caraloguing all rice genes: large -scale sequencing of randomly chosen rice cDNAs from a callus cDNA library.* The Plant J. 6(4):615-624
52. Sasaki T, M Yano, N Kurata, K Yamamoto. 1996. *The panese Rice Genome Research Program.* Genome Research 6:661-666
53. Schmidt R, J West, K Love. 1995. *Physical mapping and organization of Arabidopsis thaliana chromosome 4.* Science 270: 480-483
54. Schwartz DC, CR Canton. 1984. *Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis.* Cell 37: 67-75
55. S ral BWS, RJ Honeycutt, AG Athely, M McClelland. 1990. *Analysis of rice (Oryza sativa L.) genome using pulsed field gel electrophoresis ande rare cuting restriction endonucleases.* Plant Mol. Biol. Reports 8: 253-275
56. Song W, B Wang, GL Wang. 1995. *Construction of a rice cosmid library and identification of five cosmid clones candidatebacterial blight resistance genes, Xa-21.* Poster 218. Plant Genome III. San Diego, calif. n. 15-19, 1995.
57. Song KM, JY Suzuki, MK Slocum, PH William, TC Osborn. 1991. *A linkage map of Brassica rapa (syn. campestris) based on RFLP loci.* Theo. Appl. Genet 82: 296- 304
58. Stuber CW, SE Lincoln, DW Wolff, T Helent ris, ES Lander. 1992. *Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers.* Genetics 134: 823-839
59. Sun TP, HM Goodman, FM Ausubel. 1992. *Cloning the Arabidopsis GA1 locus by genomic subtraction.* Plant Cell 4 : 119-128
60. Tanksley SD, MW Ganai, GB Martin. 1995. *Chromosome landing: A paradigm for map-based gene cloning in plant species with large genomes.* Trends Genet 11:63-68
61. Tanksley S D, MW Ganai, JP Prince, MC De Vincent, MW Bonierbale, P Broun, TM Fulton, JJ Giovanonni, S Grandillo, GB Martin, R Messeguer, JC Miller, L Miller, AH Paterson, O Pineda, MS Roder, RA Wing, W Wu, ND Young.1992. *High density molecular linkage maps of the tomato and potato genomes.* Genetics (132): 1141- 1160
62. Umehara Y, A Inagaki, H Tanoue, Y Yasukochi, Y Nagamura, S Saji, Y Otsuki, T Fujimura, N Kurata, Y Min e. 1995. *Construction and characterization of a rice YAC library for physical mapping.* Molecular Breeding 1: 79-89

63. Wang GI, TE Holsten, WY Song. 1995. *Construction of a rice bacterial artificial chromosome library and identification of clones linked to the Xa-21 disease resistance locus*. Plant J. 7:525-533
64. Wang ML, MD Atkinson, CN Chinoy, KM Devos, RL Harcourt, CJ Liu, WJ Rogers, MD Gale. 1991. *RFLP-based genetic map of rye (Secale cereale)*. I. Chromosome 1R. Theo. Appl. Genet 82: 174-178
65. William JGK, AR Kubelic, KJ Livak, JA Rafalski, SV Tingey. 1990. *DNA polymorphism amplified by arbitrary primers are useful as genetic markers*. Nucleic Acids Res. 18: 6531-6535
66. Yang D. 1997. *Construction of a bacterial artificial chromosome library and initial physical map of rice genome*. Ph.D Thesis. Wuhan University, China.
67. Zhang HB, S Choi, SS Woo. 1995. *Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population*. Poster 276. Plant Genome III. San Diego, Calif. n. 15-19, 1995.

Thực hành

CÁCH GHI SỐ LIỆU ĐỂ CHẠY CHƯƠNG TRÌNH MAPMARKER

*Thí dụ : Cho điểm tại 33 loci của marker RG trong 47 dòng cận giao tái tổ hợp
(RIL) trên cây lúa*

NHIỆM THỂ SỐ 1

RG381	APPPP	PPAPP	PPPPP	PPPPP	PPAPP	PHPAP	PPPPP	PPPPP	PPPAP	PP
RG472	PPP-P	PPPPP	HHAPP	HAAPA	PA-PP	PAAPA	PAAAP	APPAA	PAPPP	P-
RG532	PAAPP	AAPPH	HHHPP	AAAPA	PPAPP	PPPPP	PAAAP	APPAA	PA-AP	PP

NHIỆM THỂ SỐ 2

RG120	PAA-P	AHPPH	HAPPP	HHPPH	PPPPP	P-PPP	AHHHP	HPPHH	PHPPP	PP
RG152	AAAAA	AAAPA	AAAHH	AAAPA	AAAPA	-AAAA	AAAAA	APPAA	PAPAP	AP
RG520	PPA-P	PPPP-	APAPP	PP-PA	PA-AP	APAPA	PP-AP	PPP-P	PPPPP	P-

NHIỆM THỂ SỐ 3

NHIỆM THỂ SỐ 4

RG214H	AHAAA	HPAPH	HHAPP	PPAPA	AAAAA	APAAA	HHAHA	PPPPP	PPPAP	AP
RG329S	AHAAA	HPAPH	HHAHH	PPAPA	AAAAA	AHAAA	HHAHA	PPHPP	PPPAP	AP

NHIỆM THỂ SỐ 5

RG13	HHPPP	APAPA	PPPPP	PPPPP	PPPAP	PPPAP	APPPP	PPPPP	PPPPP	PP
RG182	PHPPP	AAPAA	HAPPP	AAPPP	PPPPP	PAHHP	AAPHP	APPAA	AAPPP	PA
RG207	AAAAA	HAAPP	AAAPP	AAAPA	APAPA	PAPAP	PAAAA	APPAA	PAPAP	AP
RG435	AAAAA	AAAAA	AAAPP	AAAPA	AAAAA	-AAAA	AAAAA	APPAA	AAPAP	AA
RG573	AHPPP	APAPA	PPPPP	PPPPH	PPAHP	PPPAP	APPPP	PPPPP	PPPAP	PP

NHIỆM THỂ SỐ 6

RG123	AAA-A	AAAPP	AHAPP	AAAPA	AAAPA	AHAAA	APAAA	APP-A	PAPAP	A-
RG213	HAAAA	AAAPP	HHHPP	AAAPA	AAPPA	A-AAA	AHAAA	APPAA	PAPPP	AP
RG424	AAAAA	AAAPH	HPAPP	AA-PA	AAAP-	A--AA	APAAA	APPAA	PAPAP	AP
RG546	AAA-A	AAAPA	HPAPP	AAAPA	AA-PA	APAAA	APAAA	APP-A	PAPAP	AP
RG648	AAAAA	AAAPH	HPAPP	AAAPA	AAAPA	--AAA	APAAA	APPAA	PAPAP	A

NHIỄM THỂ SỐ 7

RG528 PPPPP PAPAP AHAPP APAPA PPPPP PAPPP PHAAP PPPPP APPPP PA

NHIỄM THỂ SỐ 8

RG333 PPAAP PPPPP PPPPP PPPPP APPPA PAPPP PPPPA PPPPP PPPPP AP

RG365 AAAAA AAAAA AAAPP AAAPA AAAPA AAAAA AAAAA APPAA PAPAP AP

NHIỄM THỂ SỐ 9

RG553 PHPPP APPPA HHPPP HPPPP PAPPP APAPA AHPHP PPPPP PPPPP PP

NHIỄM THỂ SỐ 10

RG527 AHPPA PPAP- HPAPP PPAPA AP-HA PPPAP HPAAA PPP-P PPPAP A-

NHIỄM THỂ SỐ 11

RG2H HHAHH PPPPP AHAPP PPAPA APAPA HPPHP PPAAH PPPPP PPPAP HP

RG118 HHA-P P-PAP AAAAA PP-AA PPAPP -APPP PAAAP PAA-P AAAAA P-

RG103 PHAAA PPPPP APAPP PPAPA APHPA APPPP PPAAP PPPPP PPPPP AP

RG167 HPA-P PPPPP AHAPP PPAPA AP-PA APPPP PPAAH PPP-P PPPAP P-

NHIỄM THỂ SỐ 12

RG190 AAA-P APAPP HHAPP HHAPA PAAPP PAAAA AAAAP HPP- PHPAP PP

RG235 AHAPP PAAAA PPPPP AHPPP PPAPP APPAP HPPPP HPPHH AHPAP PA

RG341 HPPPP PHAPP PPPPP HAPPP PAAPP A-AAA PPPPP APPAA PAPAP PP

RG396 AAAPP APAPP HHAPP PHAPA PA-PP AAAAA AHAAP APP-H HHPAP P-

RG463 HPPAA PPPPP PHPPP HAPPP AAAPA PPAPA PHPPA APPAA PAPAP AP

RG543 HPPAA P-PPP PHPPP HAPPP AAAPA PPAPA PHPPA APPAA PAPAP AP

Chúng ta cho điểm A : không có băng, P: có băng, và H băng ở dạng dị hợp, - : biểu thị cá thể bị mất hoặc băng không rõ ràng, sau khi chạy điện di và đọc trên phim. Xếp 5 chữ thành một nhóm. Hãy sử dụng MAPMARKER để vẽ bản đồ liên kết gen.

Chương VIII

PHẢN ỨNG CHUỖI POLYMERASE (PCR)

Phản ứng chuỗi polymerase (polymerase chain reaction) được viết tắt là PCR. Nó là công cụ khá mới trong sinh học phân tử, đánh dấu một bước tiến cực kỳ quan trọng giống như các khám phá trước đây về restriction enzyme và phương pháp Southern blot. Việc phát minh ra máy thermocycler (chu kỳ nhiệt độ được điều chỉnh một cách tự động), và sử dụng DNA polymerase có khả năng chịu đựng một phổ khá rộng về nhiệt độ như *Taq* đã giúp cho Kary Mullis đoạt giải thưởng Nobel 1993, với hàng loạt các bằng sáng chế được cấp mang số hiệu 4,683,195 (tháng 7-1987), 4,683,202 (tháng 7-1987) và hàng loạt các công bố của Saiki và ctv. (1985), Mullis và ctv. (1986), Mullis và Faloona (1987) đã minh chứng thêm thành tựu quan trọng này.

8.1. NGUYÊN TẮC CHUNG

PCR được thực hiện trên cơ sở phản ứng sinh tổng hợp DNA theo 3 trình tự như sau :

1. Biến thành dây đơn (denature)
2. Phản ứng của primer gắn vào đầu dây chuỗi mã đối xứng với chuỗi mã trên dây template (gọi là tiến trình annealing) để có phân tử DNA mới.
3. Kéo dài dây mới nhờ primer (extension)

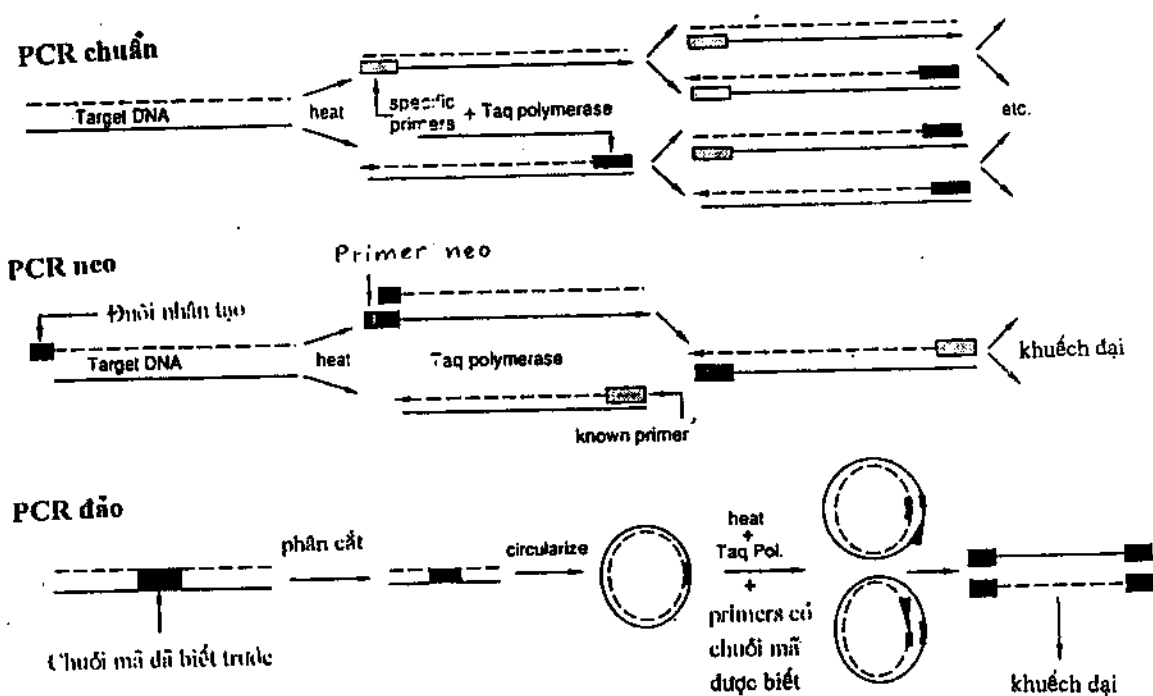
Những phản ứng này được thực hiện nhờ polymerase như *Taq* polymerase, và sự thay đổi chu kỳ nhiệt một cách hợp lý, đặc biệt là nhiệt độ cho denature và nhiệt độ cho annealing.

Ba trình tự này xảy ra theo chu kỳ rất nhanh để khuếch đại DNA(hình 4-7, 8-1).

Sự phát hiện và phân lập một DNA polymerase có tính ổn định rất cao đối với nhiệt độ, từ một vi khuẩn có tính chất chịu nhiệt (thermophilic) là *Thermus aquaticus* (*Taq*), cho phép chúng ta có thể tổng hợp dây DNA mới có tính chất lắp đi, lắp lại nhiều lần (Saiki và ctv. 1988), số lượng khuếch đại tăng theo hàm số mũ, làm cho công nghệ sinh học có thêm một phương tiện rất có giá trị cho ứng dụng của nhiều lĩnh vực nghiên cứu khác nhau.

Một polymerase của DNA và bốn deoxy (dATP, dCTP, dGTP, dTTP) được dùng để khuếch đại (amplified) một đoạn chuyên tính nào đó của DNA. Thế tăng hoạt có tính "chọn lọc" này được hoàn thiện thông qua sự hợp lại của hai oligonucleotide (F và R) trong phản ứng. Nhiệm vụ của những oligonucleotide này là tác động làm cho DNA bị biến chất (mở dây đôi thành dây đơn) và những phần cuối của đoạn nhằm tăng hoạt lên và cho nó

hoạt động như những primer trong sinh tổng hợp DNA. Để có tác động một cách đặc biệt tại vị trí mình mong muốn, những primer này phải được tổng hợp để cung cấp khoảng 20-24 bp chuỗi xoắn (duplex) tại mỗi đoạn cuối của segment.



Hình 8.1 : Phản ứng chuỗi polymerase (PCR) với 3 kỹ thuật : PCR chuẩn, PCR neo, và PCR đảo. Trong PCR chuẩn, DNA mục tiêu được mở dây đôi, sau đó có hai primer tác động ở hai đầu dây đơn, rồi DNA mới được tổng hợp nhờ Taq, với 4 deoxyribonucleotide. Trong PCR neo, chỉ cần có một primer, một đuôi dG nhân tạo được gắn vào đầu 3' của DNA mục tiêu, DNA được khuếch đại nhiều lần nhờ một primer có chuỗi mã đã biết trước, và một primer thứ hai có oligo dC. Trong PCR đảo, người ta dùng nó để khuếch đại đoạn phân tử kế cận chuỗi mã DNA đã biết trước, phân tử DNA này được phân cắt và nối lại bởi enzyme để tạo thành một vòng tròn có tính chất monomer, sau đó DNA được khuếch đại nhờ primer tương đồng với đầu của chuỗi mã được biết trước.

Primer ở bên trái tác động trên dây DNA 3'-5' còn được gọi là forward primer, ký hiệu là F. Primer bên phải tác động dây 5'-3' còn được gọi là reverse primer, ký hiệu là R. Sự sắp xếp như vậy đảm bảo vùng bị can thiệp được tăng cường hoạt động, theo 3 trình tự đã nói ở trên.

Ba giai đoạn này lặp lại nhiều lần, để tạo mức độ cao trong sự khuếch đại. Ba giai đoạn xảy ra trong điều kiện nhiệt độ khác nhau (thí dụ 94°C, 55 °C, và 72 °C), nên máy thermocycler phải có tính chất tự động hóa một cách chính xác, với sự trợ giúp của Taq. Đó là nguyên tắc cơ bản quyết định sự thành công của PCR. Sự lặp lại theo chu kỳ thường xảy ra khoảng 30 lần, để làm tăng hoạt các DNA muốn có tại vùng mục tiêu vào khoảng 100

triệu fold, đủ để tăng hoạt một gen thành một khối lớn hơn rất nhiều lần tổng số so với DNA khởi đầu. Theo nguyên tắc PCR được áp dụng để khuếch đại các đoạn DNA 200-2000 bp về chiều dài. Tóm lại, khi các primer kết hợp với sợi DNA đối lập của nó trong điều kiện một khoảng cách đã được kích hoạt, các đoạn DNA này có thể sẽ được khuếch đại lên theo phản ứng dây chuyền với polymerase. Độ dài của DNA được tăng hoạt tùy thuộc vào số đoạn và chiều dài của primer trong phản ứng, điều kiện chất đệm, nhiệt độ annealing,...

8.1.1. Qui trình chuẩn

Chu kỳ 25-35 trong điều kiện :

"Denature"	94-96 °C	15 giây (hoặc lâu hơn tùy theo vật liệu)
"Annealing"	55 °C (50-56 °C)	30 giây
"Extension"	72 °C	1,5 phút

Chu kỳ được kết thúc bằng một "extension" ở 72°C trong 5 phút. Sau đó người ta cho dừng hẳn bằng cách tồn trữ sản phẩm PCR ở 4°C.

Phức tạp nhất là nhiệt độ trong giai đoạn "annealing". Chúng ta phải điều chỉnh để có khuếch đại tốt nhất và ghi nhận được đa hình tốt nhất khi điện di trên gel (hình 8-3).

8.1.1.1. Nồng độ enzyme

Trong qui trình chuẩn, nồng độ enzyme cũng được xem xét thận trọng. Người ta khuyến cáo sử dụng 1-2,5 đơn vị (SA = 20 đơn vị / pmol) trên 100µl dung dịch phản ứng. Thông thường người ta sử dụng 0,5 - 5 đơn vị / 100µl. Nếu nồng độ *Taq* quá cao, những sản phẩm không chuyên tính có thể nhảy vào làm sai lệch kết quả. Nếu nồng độ *Taq* quá thấp, chúng ta sẽ không có đủ số lượng men để xúc tác tạo ra sản phẩm PCR theo ý muốn.

8.1.1.2. Deoxynucleotide triphosphate

Dung dịch dNTP làm stock phải được trung tính ở pH = 7,0. Stock ban đầu được pha thêm để có 10mM, và tồn trữ ở -20°C. Stock được dùng cho xét nghiệm nên có nồng độ dNTP là 1mM. Sự ổn định của dNTP trong suốt các chu kỳ lặp lại ước còn khoảng 50% sau 50 chu kỳ (Innis và ctv.1990). Hàm lượng deoxynucleotide trong khoảng 20 đến 200 µM cho kết quả ổn định, trung thực, và chuyên tính. Bốn dNTP được xử lý sao cho nồng độ chúng tương đương nhau, để giảm thiểu tối đa hiện tượng sai biệt do kết hợp sai (misincorporation) mã di truyền nào đó.

8.1.1.3. Hàm lượng Magnesium

Kết quả sẽ tốt nếu chúng ta cho vào một nồng độ Mg tối hảo cho phản ứng. Nồng độ Mg có thể ảnh hưởng đến :

- Quá trình annealing của primer.
- Nhiệt độ để tháo dây thành dây đơn.
- Sự chuyên tính của sản phẩm PCR.
- Hoạt động của enzym và sự trung thực của kết quả.

Taq yêu cầu phải có Mg tự do ở vị trí đỉnh, nơi xảy ra sự kiện gắn primer, dNTP và dây đơn DNA với nhau. Theo mục 8.1.1.2, PCR sẽ cần một hàm lượng Mg từ 0,5 đến 2,5mM ứng với hàm lượng dNTP tổng số đã cho. Ghi nhận thêm là: sự có mặt của EDTA và một vài hoạt chất bám khác (chelators) trong dung dịch stock của primer hay của dây đơn DNA sẽ có thể gây xáo trộn nồng độ tối hảo của Mg.

Tóm lại theo qui trình chuẩn này, chúng ta cần tuân thủ các qui định như sau :

Phương tiện và hóa chất :

- Máy Thermal cycler (PE 9600 Perkin-Elmer)
- Mẫu DNA ly trích 100ng / μ l
- Dung dịch stock của dNTP (mỗi dNTP là 5mM)
- *Taq* polymerase (5 U/ μ l Perkin-Elmer)
- Dung dịch stock của 10x PCR buffer(buffer của Perkin-Elmer có 15mM $MgCl_2$, 500mM KCl, 100 mM Tris-HCl, pH 8,4, 0,1% gelatin, 1% Triston X-100
- Primer F và R : 10 μ M

Chuẩn bị 25 μ l dung dịch có thuật ngữ là "cocktail" bao gồm

- | | |
|-------------------------------------|-------------|
| • DNA | 1 μ l |
| • 10x dung dịch buffer | 2,5 μ l |
| • Mỗi primer | 1 μ l |
| • Dung dịch stock dNTP | 1 μ l |
| • <i>Taq</i> polymerase | 0,1 μ l |
| • Thêm nước cất hai lần, sao cho đủ | 25 μ l |

Chu kỳ nhiệt trong thermal cycler: (Thí dụ tìm đa hình của *SacI*, clone BS3, 4kb trong người)

- 94°C trong 5 phút, theo sau là 30 chu kỳ
- 94°C trong 30 giây
- 60°C trong 30 giây
- 72°C trong 1 phút
- Giữ mẫu ở 4°C

8.2. TỐI ƯU HÓA CÁC ĐIỀU KIỆN CHO PHẢN ỨNG PCR

8.2.1. Sequence của primer

Thông thường người ta cần có primer với chuỗi mã dài cỡ 18 nucleotide trở lên, áp dụng cho genome của những sinh vật thuộc eukaryote, và primer ngắn - khoảng 10 nucleotide cho những DNA được clone hóa như cosmid. Tuy nhiên tùy theo loại primer tương hợp với mục tiêu lựa chọn marker phân tử trong xét nghiệm PCR, người ta có những primer của những marker có tính ngẫu nhiên như RAPD rất ngắn (10-16 mer), hoặc STS marker cần primer có độ dài 20-24 mer. Tuy nhiên, chúng ta nên nhớ rằng, DNA không phải là một chuỗi mã của những nucleotide một cách ngẫu nhiên, mà nó còn có đặc tính của những họ (gene families), những nguyên tố có tính lặp lại, sự lặp đoạn đơn giản, hay những chuỗi mã lặp lại phức tạp (thí dụ alpha satellite sequence của genome người). Tùy theo loài sinh vật, và tùy theo cách thể hiện của sequence dây đơn đối lập (template), người ta sẽ thiết kế chuỗi mã của primer, và kiểm tra chuỗi mã cẩn thận so với số liệu lưu trữ trước đây, nhằm loại trừ các sai sót về kỹ thuật. Gần đây, người ta thường sử dụng phần mềm trong puter để thiết kế ra những primer theo mục tiêu nghiên cứu.

8.2.2. Nhiệt độ trong quá trình annealing

Suggs và ctv. (1981) đã đề nghị một công thức để ước đoán nhiệt độ T_a , mà ở nhiệt độ ấy, một nửa các primer sẽ được phản ứng annealing với sequence mục tiêu. Công thức được sử dụng trong điều kiện primer có khoảng 20 oligonucleotide :

$$T_a = 4 (G + C) + 2 (A + T)$$

A, T, C, G là số lượng các base có trong oligonucleotide.

Tuy nhiên việc tính toán này chỉ có tính chất tạm thời, chúng ta phải điều chỉnh lại nhiệt độ này để có sự khuếch đại cao nhất và thu nhận tần suất đa hình cao nhất. Bằng kinh nghiệm, và thực hiện xét nghiệm nhiều lần, chúng ta mới có thể có số liệu đúng về nhiệt độ cho annealing. Số base càng ít, nhiệt độ này càng thấp, và ngược lại. Hiện nay có những máy thermal cycler được thiết kế cho phép chúng ta đưa vào nhiều nghiệm thức T_a để xác định nghiệm thức tối hảo.

8.2.3. Phản ứng dung dịch đệm

Một trong những ảnh hưởng có tính chất quyết định là nồng độ Mg^{2+} , kể cả về tác dụng chuyên biệt và kết quả PCR. Môi trường có tính chất ion như vậy làm cho dung dịch đệm trở nên năng động hơn rất nhiều. Do đó nồng độ Mg^{2+} cao hay thấp đều tạo ra ảnh hưởng rất khác nhau trong phản ứng PCR. Nồng độ cao của Mg^{2+} sẽ làm cho phân tử DNA dây đôi ổn định hơn, và ngăn ngừa sự biến chất hoàn toàn (sự mở dây để cho dây đơn) của sản phẩm trong mỗi chu kỳ, làm cho kết quả PCR ít đi. Lượng dư thừa Mg^{2+} còn làm cho hiện tượng annealing giả xảy ra thêm ổn định hơn, tại những vị trí không đúng của dây đơn,

cho ra những sản phẩm không mong muốn với số lượng khá lớn, nhưng mức độ chuyên tính rất thấp (Oste 1989).

Ngược lại, nếu nồng độ Mg^{2+} quá thấp ($< 0,5\mu M$), nó sẽ làm xấu đi quá trình extension (kéo dài chuỗi mã đối lập với dây gốc), trong đó Mg^{2+} đóng vai co-factor của *Taq* polymerase. Một vài ion Mg^{2+} sẽ được kẹp giữ bởi dNTP trong phản ứng. Do đó, người ta cần phải xác định nồng độ tối ưu của Mg^{2+} nhằm đảm bảo kết quả số lượng sản phẩm (yield) và sự chuyên tính của sản phẩm PCR (specificity). Trong một vài vùng có tính chất recalcitrant nghĩa là có rất nhiều cặp GC, thì chỉ cần có một khoảng biến thiên rất hẹp về nồng độ Mg^{2+} , đủ cho ra kết quả PCR và sự chuyên tính mong muốn.

Môi trường đệm KCl đã và đang được áp dụng rộng rãi, cũng có thể là buffer rất hữu dụng cho kỹ thuật PCR. Tuy nhiên, trong một vài loci, người ta thấy môi trường đệm này rất khó dùng cho khuếch đại sản phẩm, chỉ trừ Mg^{2+} đã được lựa chọn.

PCR của những loci giàu GC cho thấy: mức độ chuyên tính gia tăng khi khuếch đại trong dung dịch đệm NaCl, nhờ sự biến chất hoàn toàn (complete denature) (Holm và ctv. 1991).

Người ta còn tìm thấy dung dịch đệm ammonium sulphate - dung dịch được sử dụng rất sớm - làm giảm những sản phẩm được phát triển một cách không hoàn toàn trong PCR với *Taq*. Phương pháp này dùng để phát hiện những băng có trọng lượng phân tử thấp chạy trên acrylamide gel, đối với những sản phẩm PCR được theo dõi nhờ đánh dấu phóng xạ.

pH của dung dịch đệm cũng được xem xét trong kỹ thuật PCR. Hầu hết các dung môi phản ứng được đệm trong môi trường 10-50 mM Tris - HCl ở pH = 8,3, nhiệt độ 20°C. Tuy nhiên, pH sẽ giảm khi nhiệt độ tăng. Theo chu kỳ nhiệt độ của PCR, pH sẽ thay đổi từ 7,8 đến 6,8. Với khoảng giá trị như vậy, pH được xem như tạm chấp nhận.

8.2.4. Nồng độ deoxynucleotide

Người ta đã khuyến cáo rằng mỗi deoxynucleotide được sử dụng là 200 μM . Lượng thừa đầu tiên của deoxynucleotide tương xứng theo 100 ng DNA của dây đơn (10^{10} đến 10^{11}), nồng độ của nó phải là một hằng số không đổi (constant). Nồng độ ban đầu là 200 μM , nồng độ dNTP sau cùng sẽ hình thành một lượng thừa 10^4 , ngay cả khi 50% của nó bị phân hủy trong quá trình diễn biến chu kỳ nhiệt. Điều quan trọng là chúng ta phải giữ cho nồng độ của tất cả các deoxynucleotide bằng nhau, để tránh sự kết gán vào nhầm lẫn.

8.2.5. DNA polymerase chịu nhiệt

Nếu DNA polymerase quá dư thừa, chúng ta sẽ thấy hiện tượng tổng hợp DNA do phản ứng giả của primer trên dây đơn. Trong hầu hết các protocol, người ta khuyến cáo nồng độ *Taq* polymerase nên sử dụng là 0,5 U / 25 μl , sẽ cung cấp enzym có nồng độ phân tử thấp nhất cho mọi hợp phần của phản ứng, 1nM. Sau khi khuếch đại được 10^6 -fold ở vùng mục tiêu khởi đầu, 1fM, chúng ta sẽ có một số lượng phân tử tương đương với phân tử

enzym. Nồng độ ít này đủ để kiểm soát sự chuyên tính của phản ứng khuếch đại. Chúng ta không nên sử dụng nồng độ cao hơn 2,5 nM của enzym (1,25 U / 25 μ l).

8.2.6. Chất ổn định hoạt động enzyme

Những hoạt chất ổn định enzym (enzyme stabilizers) cũng được chú ý. Nhiều nhà sản xuất hóa chất đã xem xét gelatin hoặc Triston X-100 trong những buffer để tồn trữ enzym. Có trường hợp người ta sử dụng cả hai làm dung dịch đệm. Nhằm bảo quản và ổn định enzym trong quá trình xử lý nhiệt của PCR, mỗi dung dịch đệm trong phản ứng đều có chứa gelatin ở nồng độ 0,01% và Triston X-100 ở nồng độ 0,1% (v/v).

8.2.7. Nồng độ các primer

Lượng thừa đầu tiên của primer ứng với dây đơn template (10^7) là nồng độ primer cần thiết, nó phải là hằng số. 95% các phân tử đầu tiên của primer vẫn còn duy trì sau 30 chu kỳ. Trong hầu hết các ứng dụng PCR, hai primer F và R đều phải có nồng độ bằng nhau. Nồng độ sau cùng của mỗi primer là 0,1 μ M, tương đương với 2,5 pmol (16,25 ng của 20-mer) trong 25 μ l dung dịch cocktail. Nồng độ này phải được điều chỉnh 0,5 μ M, tùy thuộc vào locus nào đó cần được khuếch đại. Những protocol trước đây khuyến cáo primer ở nồng độ 3 μ M, nhưng với nồng độ cao như vậy không những không cần thiết mà còn tạo ra bất lợi, vì quá thừa primer, dẫn đến sự hình thành hiện tượng primer dimers, và hiện tượng priming nhầm.

8.2.8. Số lượng DNA

Phân tử DNA ở đây hiểu theo nghĩa DNA template.

Thí dụ genome của người, DNA có 125 ng / 25 μ l được sử dụng. Điều này tương đương với 0,063 attomoles của DNA ($\# 10^{-19}$ mol) và ở nồng độ 2,5 fM. Khi khuếch đại vùng mục tiêu từ dây đơn DNA, số lượng của template được tính toán trên cơ sở mol. Phép tính này giúp chúng ta xác định số lượng DNA cần phải có, số chu kỳ để tổng hợp ra DNA với số lượng mong muốn.

Tính trung thực (fidelity) của việc khuếch đại nhiều ít tùy thuộc vào mức độ sai sót của DNA polymerase. Tính trung thực được hiểu là số copy được tổng hợp nhờ khuếch đại theo chu kỳ PCR sẽ giống hệt như phân tử DNA gốc. Nó còn tùy thuộc vào số copy đầu tiên của template. Sai số ngẫu nhiên có trong vài chu kỳ đầu sẽ lan rộng qua những chu kỳ sau. Tình trạng này được xem xét qua hiện tượng jackpot - một kiểu đột biến được tìm thấy trên *E. coli*. Nếu số lượng phân tử khởi đầu ít hơn 100, và tính trung thực của phản ứng polymer được giữ tốt, thì người ta có thể thực hiện những chu kỳ đầu tiên của PCR, kiểm soát được bằng T7 hoặc T4 polymerase. Cho dù những enzym này không có tính chịu nhiệt giới, nhưng nó cũng sẽ bổ sung thêm sau mỗi lần denature (mở dây đơn). Nó cho mức độ sai sót thấp hơn *Taq* polymerase (Keohavong và Thilly 1989). Sau một vài chu kỳ đầu tiên, PCR tự điều chỉnh lại với hoạt động của *Taq* polymerase.

Bảng 8.1 : Những polymerase chịu nhiệt và các điều kiện kèm theo

Enzym	Buffer / pH	Muối	Cation Mg^{2+}
<i>Taq</i> <i>Thermus aquaticus</i>	10mM Tris-HCl, pH 8,3	50mM KCl	1,5-5,0mM $MgCl_2$
<i>Ultma</i> <i>Thermotoga maritima</i>	10mM Tris-HCl, pH 8,8	10mM KCl	1,5-5,0 mM $MgCl_2$
<i>Tth</i> <i>Thermus thermophilus</i>	10mM Tris-HCl, pH 8,3	90mM KCl	1,0-2,0 mM $MgCl_2$ +reversetranscriptase 2,0-4,0 mM $MgCl_2$ + EGTA để giữ Mg^{++}
<i>Pfu</i> <i>Pyrococcus furiosus</i>	20mM Tris-HCl, pH 8,2	10mM KCl 6mM (NH_4) ₂ SO ₄	1,2-2,5 mM $MgCl_2$
<i>Tli</i> <i>Thermococcus littoralis</i>	10mM Tris-HCl, pH 9,0	50mM KCl	1,5-5,0 mM $MgCl_2$
<i>Tub</i> <i>Thermus ubiquitus</i>	50mM Tris-HCl, pH 9,0	20mM (NH_4) ₂ SO ₄	0,7-2,0 mM $MgCl_2$
<i>Tfl</i> <i>Thermus flavus</i>	50mM Tris-HCl, pH 9,0 hoặc 20mM Tris-acetate, pH 9,0	20mM (NH_4) ₂ SO ₄ 70mM K-acetate	1,5-5,0 mM $MgCl_2$ tổng hợp DNA 1,5-5,0 mM $MnSO_4$ reverse transcriptase
<i>Pwo</i> <i>Pyrococcus woesei</i>	10mM Tris-HCl, pH 8,85	20mM (NH_4) ₂ SO ₄	1,5-4,0 mM $MgSO_4$
<i>Tbr</i> <i>Thermus brokianus</i>	10mM Tris-HCl, pH 8,8	50mM KCl	1,5-5,0 mM $MgCl_2$

Sự khuếch đại của những mẫu DNA như vậy (≤ 100 copy của dây đơn template) được công bố như quá trình có hai pha: pha thanh lọc và pha khuếch đại. Pha thanh lọc có đặc điểm là một lượng dư thừa cao các hóa chất ứng với một ít phân tử của dây template. Phân tử này được pha loãng ra, làm cho giai đoạn thanh lọc của PCR trở nên khó khăn hơn, vì tần suất để primer và dây template gặp nhau giảm rõ rệt, mà tiếp xúc giữa primer và primer lại tăng lên. Tình trạng này hình thành hiện tượng primer-dimers. Phương pháp booster PCR (kích hoạt) đã được đề nghị để khắc phục hiện tượng này. Theo tiến trình khuếch đại vùng mục tiêu gấp nhiều lần (10^7 fold) trong suốt pha thanh lọc, những phân tử được tạo ra đầu tiên sẽ trở thành dây template chính cho lần khuếch đại tiếp theo. Phương pháp booster PCR đã thành công trong việc khuếch đại những phân tử đơn DNA, với kết quả sản phẩm PCR cao, với rất ít hiện tượng primer-dimers, hơn phương pháp PCR tiêu chuẩn. (Ruano và ctv. 1991).

Phương pháp hot-start PCR (khởi động nóng) trở nên phổ biến trong trường hợp chẩn đoán và pháp y để khuếch đại một mẫu DNA rất loãng, nghĩa là số lượng DNA cực kỳ thấp. Tất cả những vật liệu trong kỹ thuật này được chuẩn bị theo qui cách đặc biệt như: dNTP, Mg^{2+} , *Taq*, primer). Ảnh hưởng nổi bật của phương pháp "hot-start PCR" là cho DNA template vào quá trình khuếch đại có chọn lọc, trong chu kỳ thứ nhất của PCR. Phương pháp hot-start PCR loại trừ bất cứ sự tổng hợp các sản phẩm ngoại sinh nào đó, do primer có tính chất không chọn lọc lại với DNA template, ở nhiệt độ $30^{\circ}C - 60^{\circ}C$, mà *Taq* polymerase duy trì hoạt động của nó ở nhiệt độ ấy. Sản phẩm giả như vậy sẽ được tổng hợp trong chu kỳ thứ nhất của PCR, nếu toàn bộ dung dịch phản ứng được chuẩn bị trong điều kiện nhiệt độ bình thường của phòng thí nghiệm, rồi xử lý ở $94^{\circ}C$. Bắt đầu vào chu kỳ thứ hai của PCR, những sản phẩm giả như vậy có thể được khuếch đại làm sai lệch mục tiêu mong muốn.

8.2.9. Tỷ lệ giữa primer : template

Một trong những khái niệm quan trọng nhất của PCR là tỷ lệ tối hảo giữa primer và template. Nếu tỷ lệ này quá cao, hiện tượng primer-dimers sẽ xuất hiện, giống như trường hợp mẫu DNA quá loãng. Nếu tỷ lệ này quá nhỏ, kết quả sản phẩm PCR sẽ không nhiều.

Hầu hết các trường hợp áp dụng PCR, đều dùng một nồng độ primer không quá $0,5 \mu M$ ($12.5 \text{ pmol}/25 \mu l$ cocktail), không tính đến nồng độ template DNA, để tránh hiện tượng primer-dimers.

8.3. ĐỊNH LƯỢNG DNA VÀ RNA BẰNG PCR

Khả năng định lượng số copy thấp của một nucleic acid nào đó rất quan trọng cho nghiên cứu cơ bản và ứng dụng thuộc lĩnh vực y học. Kết hợp hai phương pháp PCR và TGGE (temperature gradient gel electrophoresis = điện di trên gel theo gradient của nhiệt độ) cho phép chúng ta một hệ thống đáng tin cậy để định lượng.

Hệ thống PCR/TGGE yêu cầu các tiêu chuẩn đồng nhất ở bên trong, ngoại trừ nucleotide, với mục tiêu xác thực rõ ràng. Các vị trí phản ứng priming đồng nhất và những chuỗi mã đồng nhất, tỉ lệ giữa thực tế và lý thuyết tương đối ổn định trong suốt quá trình khuếch đại. Sau khi phản ứng biến chất và hoàn nguyên dây DNA, các phân tử có đánh dấu hình thành dây đôi dạng homo với khuếch đại chuẩn (amplified standard), và hình thành dạng hetero với khuếch đại mục tiêu (amplified target). Liên sau đó, phương pháp TGGE sẽ giúp chúng ta phân loại hai dạng homoduplex và hetero duplex này.

Dạng homo di chuyển nhanh hơn dạng hetero trong những lần mở dây đơn cuối cùng. Số lượng copy đầu tiên của nucleic acid mục tiêu rất dễ tính, bằng cách lấy theo tỉ lệ giữa hetero và homoduplex, và số copy đầu tiên của nucleic acid chuẩn. Đối với định lượng RNA, một RNA tiêu chuẩn sẽ được dùng làm kiểm chứng hiệu quả của tổng hợp cDNA đối với sự khuếch đại PCR.

Ngoài ra, hệ thống PCR/TGGE còn phát hiện những thể đột biến (target mutants)

Trình tự của hệ thống này được thực hiện như sau :

- Phân tích TGGE.
- Định lượng DNA bao gồm: tổng hợp DNA chuẩn, xác định số copy của plasmid, xác định số copy của DNA trong mẫu.
- Định lượng RNA : tổng hợp RNA chuẩn, định lượng mRNA.

Trong phân tích TGGE, chúng ta phải chuẩn bị polyacrylamide gel bằng cách pha trộn.

- | | |
|------------------------|--------|
| • urê | 21,6g |
| • 30% acrylamide stock | 12ml |
| • 1M MOPS | 0,9ml |
| • 0,5 M EDTA | 90µl |
| • 40% glycerol | 2,25ml |

Sau đó chuẩn bị mẫu PCR :

- | | |
|--------------------------|-----|
| • sản phẩm PCR | 4µl |
| • DNA chuẩn có đánh dấu | 1µl |
| • 2X dung dịch đệm (TAE) | 5µl |

Cho vào heat ở 98°C, 5 phút để phản ứng biến chất xảy ra, rồi hoàn nguyên ở 50°C trong 15 phút. Chạy điện di TGGE (trên dung dịch đệm 1X TAE hay TE), ở hiệu thế 300V, thời gian 2 giờ. Phấm màu xanh bromophenol chuyển dịch 7-9cm dọc theo mỗi slot.

Giữ yên gel trong tủ định ôn, có 10% ethanol, 0,5% acetic acid trong vòng 15 phút.

Định ôn gel có 2% glycerol trong 10 phút và sấy khô 50°C trong 3 giờ.

Chụp gel trên phim X-quang, và đánh giá kết quả trên phim.

8.4. THIẾT KẾ PRIMER

Việc chọn lọc một primer nào đó là một thử thách cực kỳ quan trọng khi chúng ta muốn ứng dụng PCR có hiệu quả và độ trung thực cao.

Nhằm đạt được thành công trong việc khuếch đại của PCR, chúng ta phải có một thiết kế chính xác về oligonucleotide primer.

Sequence của primer được chọn xác định kích thước, vị trí của sản phẩm PCR, cũng như nhiệt độ T_m vùng được khuếch đại. Primer được thiết kế đúng có thể giúp chúng ta tránh tạo ra những sản phẩm PCR không chuyên tính, giống như hiện tượng phân biệt giữa cDNA và DNA của genome trong RNA-PCR.

8.4.1. Thông số cơ bản để thiết kế primer

Mục đích của thiết kế primer là có được một cân bằng giữa hai kết quả: sự chuyên tính và sự hiệu quả của PCR. Sự chuyên tính được xem xét trên cơ sở xuất hiện bao nhiêu lần hiện tượng priming nhằm trên tổng số lần priming (sự lai giữa primer và dây template tại vị trí mục tiêu của nó). Tính hiệu quả của một primer là làm thế nào tiếp cận với lý thuyết về kết quả sản phẩm, trong mỗi chu kỳ PCR, một cặp primer có thể khuếch đại ra một sản phẩm.

Một chuỗi mã DNA nào đó đã được cung cấp, người ta có thể phân tích primer trên phần mềm máy tính, để có một cân bằng giữa hai mục tiêu nói trên..

8.4.1.1. Chiều dài primer

Sự chuyên tính thường được điều khiển bởi chiều dài của primer và nhiệt độ trong quá trình annealing. Oligonucleotide có kích thước 18-24 base có xu thế trở thành những sequence rất chuyên tính, nếu nhiệt độ annealing của PCR được thiết kế xê xích vài độ so với T_m của primer (nhiệt độ lý thuyết). Primer có chiều dài càng lớn, sự chia các template để được quán đôi với primer - càng nhỏ. Trong khi khuếch đại theo hàm số mũ, một sự cố nhỏ nào đó của quá trình annealing cũng sẽ làm suy giảm kết quả sản phẩm PCR.

Để tối ưu hóa PCR, người ta sử dụng primer có chiều dài tối thiểu sao cho đảm bảo nhiệt độ melting khoảng 54°C, hoặc hơn chút ít, điều này sẽ cung cấp những cơ hội tốt nhất để duy trì sự chuyên tính và mức độ hiệu quả cao.

Những oligonucleotide ngắn (15 base hoặc ngắn hơn) được dùng một cách hạn chế trong nhiều qui trình PCR, thí dụ như primer ngẫu nhiên trong mapping các genome đơn

giản (Liang và Pardee 1992, William 1990). Tùy thuộc vào kích thước genome của sinh vật, chúng ta sẽ có một độ dài tối thiểu, sai biệt hơn kém nhau vài nucleotide. Chiều dài primer tối thiểu cho hầu hết các ứng dụng PCR là 18 nucleotide.

Nếu cDNA thuần khiết được sử dụng, hoặc nếu DNA của genome không có, thì chiều dài này có thể giảm xuống, để mức độ rủi ro do hiện tượng tương tác không chuyên tính giữa primer và dây template sẽ giảm. Người ta thường thiết kế primer có độ dài 18-24 mer.

Mặt khác, primer càng ngắn, quá trình quán đôi giữa primer và dây DNA càng nhanh, tạo thành dây đôi ổn định trong tổng hợp DNA.

Thông thường, primer có 28-35 mer cần cho việc khuếch đại các sequence có mức độ dị hợp cao.

Có hai khả năng áp dụng trong thiết kế chiều dài của primer :

- Khuếch đại những phân tử có tính đồng dạng (isoform) của protein hoặc một họ protein trong cùng một loài sinh vật, cũng như cloning một gen của loài khác mà sequence của nó rất cần thiết cho nhà nghiên cứu (Dveksler và ctv. 1993)
- Khuếch đại những sequence của virus như HIV-1, mà sự biến thiên của sequence, sự có mặt của sequence, sẽ xác nhận sự có mặt của bệnh do virus gây ra (Mack và Sninsky 1988, Ou và ctv. 1988)

Trong cả hai trường hợp, người ta đều nhờ phần mềm của máy tính để thiết kế primer, nhằm so sánh tất cả các sequence có liên quan, và mô tả vùng của DNA mục tiêu, với số lượng thấp nhất sự biến thiên của chuỗi mã.

Khi chọn lựa primer để khuếch đại DNA từ những loài sinh vật khác nhau, chúng ta nên tránh chuỗi mã ở vùng không giải mã được 5'-3' của mRNA, bởi vì nó có thể không có bất cứ một mức độ tương đồng nào (homology).

Vị trí đầu 3' của primer nên được xác định cẩn thận, vì nó có tính chất quyết định cho sự thành công của PCR. Khi chúng ta xác định được một amino acid, hai base đầu tiên trong codon của nó, hoặc 3 base trong trường hợp nó được mã hóa bằng codon đơn [methionine và tryptophan], thì 2 hoặc 3 base đầu tiên này được dùng như đầu 3'.

Những cặp base hoàn chỉnh giữa những đầu 3' của primer và template cho phép chúng ta có được kết quả mong muốn, và giảm thiểu tối đa hiện tượng không tương xứng (mismatches) trong khoảng 5-6 nucleotide tại đầu 3' của primer.

Thông thường, sự thêm vào sequence không có liên quan tại đầu 5' của primer vẫn không làm thay đổi quá trình annealing. Trong một vài trường hợp, khi số lượng base nhiều lên có ý nghĩa, những base này không tương xứng với sequence của dây template, được

thêm vào primer. Sự khuếch đại với 4 - 5 chu kỳ sẽ có thể hoàn tất, trong điều kiện nhiệt độ annealing thấp.

Đối với những primer ngắn hơn 20 mer, Sugg và ctv. (1981) đã đề xuất một công thức tính T_m liên hệ với các base (xem mục 8.2.2). Trong khi primer dài hơn, tính toán này chỉ có giá trị gần đúng. Sau này nhờ phương tiện computer, Breslauer và ctv. (1986), Freier và ctv. (1986) đã thiết kế PCR primer theo chương trình cài đặt sẵn, khá tiện lợi.

8.4.1.2. Nucleotide cuối cùng của primer

Nhiệm vụ của đầu 3' trong primer là kiểm soát hiện tượng priming nhầm (Kwok và ctv. 1990). Nhiệm vụ khác của nó là ngăn ngừa hiện tượng tương đồng (homology) trong cùng một cặp primer. Chúng ta phải thận trọng vì primer không phải là sự kiện cái này bổ sung cho cái kia, đặc biệt tại đầu 3' của nó. Nếu không thận trọng trong việc xác định đầu 3', hiện tượng primer-dimers sẽ xảy ra, với tính chất bổ sung cho nhau, sản phẩm PCR lúc bấy giờ sẽ là một sự khuếch đại của chính primer, chứ không phải DNA template mà ta mong muốn. Trong trường hợp có nhiều cặp primer đưa vào trong cùng một phản ứng (multiplex PCR), chúng ta cần phải kiểm tra gấp đôi hiện tượng bổ sung cho nhau của tất cả primer. Thông thường, trong chương trình cài đặt sẵn của computer, không cho phép cặp primer có sự tương đồng về đầu 3', nhằm giảm cơ hội xuất hiện hiện tượng primer-dimers (Chou và ctv. 1992).

8.4.1.3. Hàm lượng GC và T_m

Primer của PCR phải duy trì được hàm lượng GC đến mức có thể được. Những oligonucleotide có 20 base, với 50% G+C, thường có giá trị T_m trong khoảng 56-62°C, tạo ra điều kiện để quá trình annealing đạt kết quả tốt. Hàm lượng GC và T_m phải khớp với một cặp primer sẽ được thiết kế. Nếu giá trị T_m càng lớn, cơ hội cho priming nhầm càng cao.

Do đó khi thiết kế primer, chúng ta phải lưu ý đến hàm lượng GC, đôi khi do làm nhiều mẫu, chúng ta sẽ có những kinh nghiệm riêng trong thiết kế primer, nhất là việc chuyển đổi sequence của RFLP sang STS marker.

8.4.1.4. Chọn lọc không dựa trên computer

Primer phải được chọn lọc từ những vùng được mô tả rất kỹ ở đầu 3' và đầu 5' của một sequence rất chuyên tính nào đó. Phương pháp thiết kế primer đơn giản ở đây là: chọn những vùng có thiếu một nucleotide đơn nào đó. Phương pháp chọn lọc primer như vậy là cách làm giảm cơ hội xảy ra hiện tượng tương đồng primer-primer. Một lần nữa, chúng ta phải thận trọng cân nhắc về chiều dài cặp primer, hàm lượng base, để giá trị T_m của primer xê xích trong vòng 2-3°C.

8.4.2. Sử dụng phần mềm để thiết kế primer

Công thức đầu tiên của Suggs và ctv. (1981) là $T_m = 2^{\circ}\text{C} \times (A+T) + 4^{\circ}\text{C} \times (G+C)$ đã trở nên phổ biến nhờ tính đơn giản và tính chính xác của nó.

Hai chương trình sử dụng hơi khác nhau một chút về cách tính toán, nhưng cùng cho ra một primer được chọn, nếu chúng ta nhập vào những thông số cơ bản giống nhau. Những sai biệt do phép tính khác nhau dẫn đến thứ tự ưu tiên trong tiêu chuẩn chọn lọc đã và đang được áp dụng.

Người ta thực hiện các phương án để dự đoán giá trị T_m trên cơ sở các thông số nhiệt động học. Trong chương trình khác, người ta mô hình hóa nhiệt độ annealing dựa trên công thức được phát triển từ các đoạn phân tử DNA có độ dài hơn 100 nucleotide (McConaughy và ctv. 1969). Công trình nghiên cứu của Rychlik và ctv. (1990) cho phép chúng ta mô hình hóa về nhiệt độ annealing tối hảo của một cặp primer. Sau đó, Wu và ctv. (1991) đã dựa trên chiều dài primer, hàm lượng GC, để mô hình hóa nhiệt độ tối hảo này. Như vậy, hầu hết các chương trình đều nhằm tìm một nhiệt độ annealing thích hợp cho PCR. Ngày càng có nhiều chương trình cải tiến với độ chính xác càng cao, đã được nhiều phòng thí nghiệm ứng dụng.

8.5. NHỮNG ỨNG DỤNG QUAN TRỌNG CỦA PCR

Hiện nay, thành tựu của PCR mở ra nhiều triển vọng cho sinh học phân tử, với nhiều ứng dụng trong sinh học, trong y khoa, trong nông nghiệp... (hình 8-2)

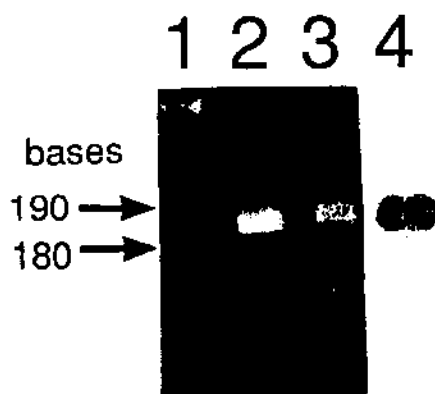
Trong nghiên cứu genome :

- Nhân bản vô tính với PCR
- Recombinant PCR
- Kỹ thuật footprinting DNase I
- Thanh lọc λ gt11 libraries
- HLA DNA
- Multiplex PCR

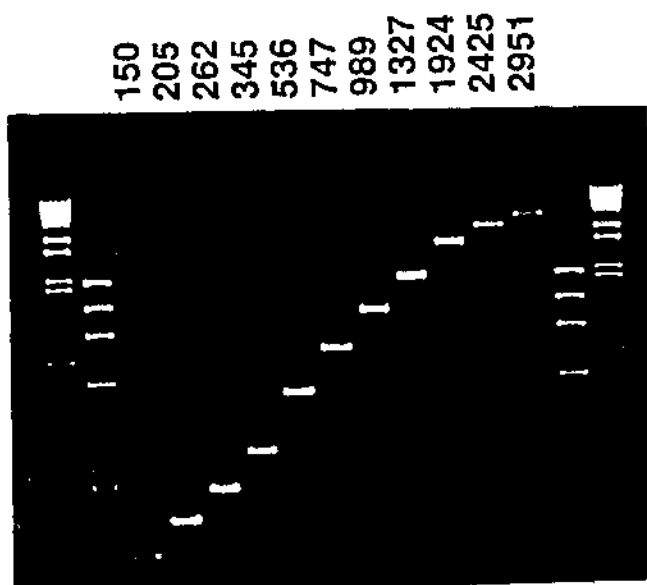
Trong y khoa :

- Phát hiện tế bào T / virus gây bệnh ung thư máu
- Phát hiện siêu vi gan B
- Nhiều ứng dụng khác trong chẩn đoán bệnh

Ở đây chúng ta chỉ đề cập đến vài ứng dụng trong phân tích genome.



Hình 8.2 : Ứng dụng PCR trong chẩn đoán bệnh ở người. Phát hiện siêu vi gan B. Cột 1, marker pBR322 được cắt bởi *MspI*. Cột 2, sản phẩm PCR chưa được biotyl hóa. Cột 3, sản phẩm PCR đã được biotyl hóa. Cột 4, phân tích Southern sản phẩm PCR đã được methyl hóa trên gel có alkaline để tạo phản ứng denature (Innis và ctv. 1990)



Hình 8.3 : Khuếch đại đoạn phân tử β -globin có kích thước từ 150 - 2951 bp. Dung tích mẫu 100 μ l bao gồm dung dịch đệm, 200 μ M cho mỗi dNTP, 250 nM cho mỗi primer, 100 ng DNA của người, 2,5 đơn vị Taq polymerase. Phản ứng xảy ra trong 30 chu kỳ. Điện di trên gel với hàm lượng agarose là 1,6% và nhuộm màu với ethidium bromide fluorescence. Chiều dài của sản phẩm thể hiện thành một vạch của mỗi cột. Marker là phage λ được phân cắt bởi *BstEII* (500 ng) và Φ x 174RI được phân cắt bởi *HaeIII* (250 ngg) (Innis và ctv. 1990)

8.5.1. Sử dụng PCR để tạo nhanh các gen tổng hợp

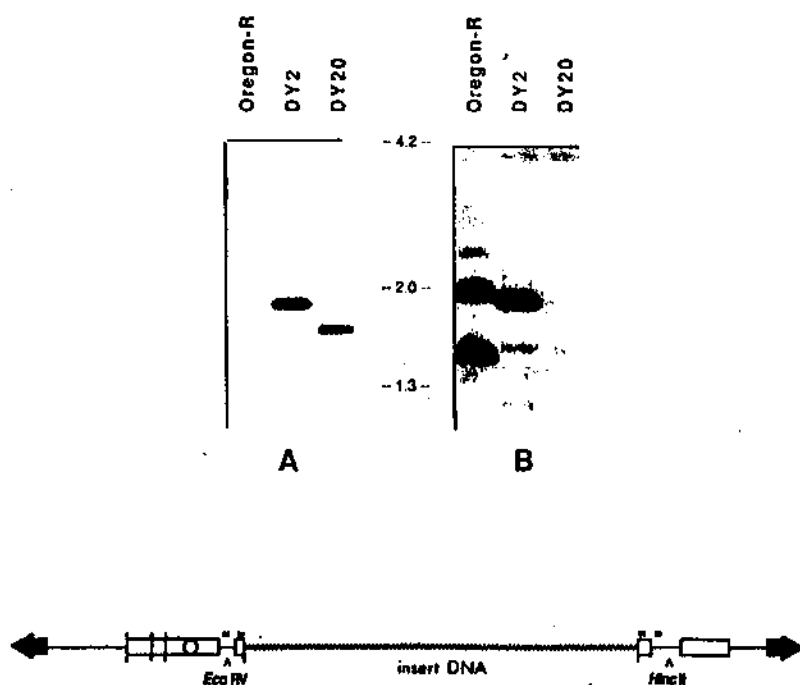
Drillon và Rosen (1993) đã khai thác khả năng sáng tạo những đoạn phân tử DNA mới của PCR. Phương pháp của họ có tên là PCR hai giai đoạn (two-step PCR) để tạo nhanh các gen tổng hợp. Phương pháp được xây dựng trên những quan sát của Mullis và ctv. (1986) trong đó những nucleotide chồng lấp nhau có thể được sử dụng để sáng tạo ra DNA tổng hợp, thông qua nhiều vòng khuếch đại PCR theo kiểu Klenow. Phương pháp này tỏ ra hữu dụng khi tổng hợp gen cho protein của HIV-2 Rev và protein Wilms - locus gây ra khối u. Ngoài ra phương pháp này còn thành công trong việc tạo ra đột biến nguyên tố phản ứng với HIV-1 rev. Khả năng tạo ra DNA tổng hợp có nhiều áp dụng tùy thuộc vào thiết kế của chúng ta. Các gen của sinh vật eukaryote thường thể hiện rất kém trong hệ thống thể hiện của vi khuẩn. Lý do của việc này là vi khuẩn có những codon của gen của sinh vật eukaryote. Người ta có thể thiết kế một gen tổng hợp có chứa codon, mà codon này được dùng trong sinh vật thường được thể hiện protein. Phương pháp này được hoàn tất nhờ sự giải mã ngược chuỗi mã amino acid của protein mong muốn. Để thay đổi cấu trúc của codon, gen tổng hợp có thể được thiết kế có những restriction site thuận lợi phục vụ cho thí nghiệm cloning, hoặc không có restriction sites chuyên tính làm phương tiện cho cloning theo chiều sâu :

1. Phát sinh thể khảm (chimeric) có tính đồng nhất phục vụ cho nghiên cứu tương tác giữa cấu trúc và chức năng, nghiên cứu ảnh hưởng trao đổi của một loại protein có và không có liên quan đến thể khảm.
2. Thay đổi trên qui mô lớn của protein hoặc những nguyên tố chuyển mã (như promoter, terminator,...) [có thể phân tích đột biến].
3. Sáng tạo ra những promoter mới hoặc protein mới.
4. Bảo hòa tính phát sinh đột biến của gen thông qua sử dụng nucleotide có tính chất ngẫu nhiên hoặc sử dụng deoxyinosine trong thiết kế sequence của gen.

Nguyên tắc của PCR hai giai đoạn có thể được mô tả như sau: hai phản ứng PCR xảy ra liên nhau. (i) phản ứng thứ nhất của PCR phát sinh ra dây template tương ứng với gen tổng hợp, (ii) nó được khuếch đại trong phản ứng PCR lần thứ hai. Trước khi bắt đầu tiến hành này, người ta phải thiết kế cấu trúc và phải xác định chuỗi mã nucleic acid của gen tổng hợp mà mình mong muốn. Sau đó oligonucleotide theo chiều dài của gen phải được thiết kế và được tổng hợp. Thông thường người ta tổng hợp oligonucleotide theo số chữ với chiều dài khoảng 16 đến 30 nucleotide. Số lượng và độ dài của từng oligonucleotide tùy thuộc vào kích thước của DNA tổng hợp, thông thường oligonucleotide có chiều dài nằm giữa 60-125 nucleotide. Thí dụ, người ta cần có 4 oligonucleotide sử dụng để tổng hợp DNA có 325 bp, và 8 oligonucleotide để tổng hợp 765 bp.

Vật liệu thường dùng trong phương pháp này là :

- 10X PCR buffer: 500mM KCl, 100 mM Tris-HCl, pH 8,0, 15mM MgCl₂.
- 10X dNTP: 2mM cho mỗi dATP, dCTP, dGTP, dTTP.
- *Taq* DNA polymerase.
- Nước cất hai lần.
- Dầu công nghiệp tinh khiết.
- Các oligonucleotide chồng lấp nhau theo chiều dài của đoạn phân tử DNA. Số chẵn của oligonucleotide nên được sử dụng, chồng lên ít nhất 15 nucleotide.
- Những primer kế cận nhau có những restriction site sẵn sàng cho cloning.
- Phân tích trên agarose gel những sản phẩm PCR.



Hình 8.4 : Khuếch đại chuỗi mã kế cận bằng PCR đảo. Tạo ra thể thăm dò từ YAC có tính chất lai ở một đầu chuyên biệt. *EcoRV* và *HincII* có vị trí xác định rất gần vị trí cloning của vectơ. Vị trí của những oligonucleotide primer và phương hướng tổng hợp DNA 9 trong dấu ngoặc kép (" ") là ưu điểm của YAC. Những vùng của vectơ YAC và pBR322 được biểu thị bằng hình hộp và một gạch ngắn, theo thứ tự. Cột 1 là Oregon-R (10 microgram) của genome ruồi giấm. Cột 2, DNA của men (1 microgram) là YAC gắn với 150 kb DNA của ruồi giấm (DY2) dùng để tạo probe có dấu đặc biệt. Cột 3, DNA của men (1 microgram) chứa YAC gắn theo mình nó DNA của ruồi giấm khác (DY20) được phân cắt nhờ *EcoRV* và *HincII*. [A] mẫu được đánh dấu bằng ³²P. [B] mẫu giống nhau có probe với dấu đặc biệt là sản phẩm PCR đảo từ DY2. Đơn vị là kilobases (Innis và ctv. 1990).

8.5.2. Kỹ thuật cloning cDNA bằng PCR đảo

Kể từ báo cáo đầu tiên về cDNA cloning (Verma và ctv. 1972), kỹ thuật này đã và đang phát triển với nhiều thành tựu nổi bật, làm công cụ rất phổ biến cho công tác phân lập gen, xác định tính chất gen, và phân tích genome cho cả sinh vật eukaryotic và prokaryotic. Kỹ thuật cloning cổ điển đối với cDNA đòi hỏi rất nhiều công sức về kho lưu trữ (library để bảo quản thực khuẩn thể, hoặc plasmid, nó cần một khối lượng công việc điều tra một số lượng lớn các phage hoặc plasmid có tính chất recombinant. Có 3 hạn chế trong phương pháp này là :

- Cần có một lượng cơ chất (ít nhất 1 μ g) mRNA tinh khiết làm vật liệu khởi đầu để tạo ra kho lưu trữ (library) với mức độ đa dạng đầy đủ cho nghiên cứu.
- Khó khăn thuộc về bản chất bên trong của những phản ứng enzyme tiếp nối nhau làm cho cDNA cloning có kết quả thấp, và những dòng vô tính (clone) bị đứt đoạn.
- Việc thanh lọc một "library" với kỹ thuật lai đòi hỏi rất nhiều thời gian.

Kỹ thuật PCR hiện nay có thể giúp chúng ta khắc phục những nhược điểm nói trên, làm cho cDNA cloning trở nên dễ dàng hơn, có những cải tiến tốt hơn (Huang và ctv. 1993).

Sử dụng hai primer có tính chuyên tính đối với gen, một cDNA có sequence đã được biết rõ, người ta cho khuếch đại trong PCR. Tuy nhiên, cái khó là làm sao phân lập những copy của cDNA toàn bộ chiều dài, từ mRNA, trên cơ sở thông tin về chuỗi mã rất hạn chế. Chuỗi mã chưa được biết này nằm kề một đoạn nhỏ của chuỗi mã đã biết rõ, chuỗi mã chưa được biết không thể được khuếch đại bằng phương pháp PCR thông thường (hình 8-4).

Gần đây, phương pháp PCR neo (anchored-PCR) và phương pháp PCR đảo (inverse PCR) đã được phát triển để giải quyết vấn đề này.

8.5.2.1. PCR neo

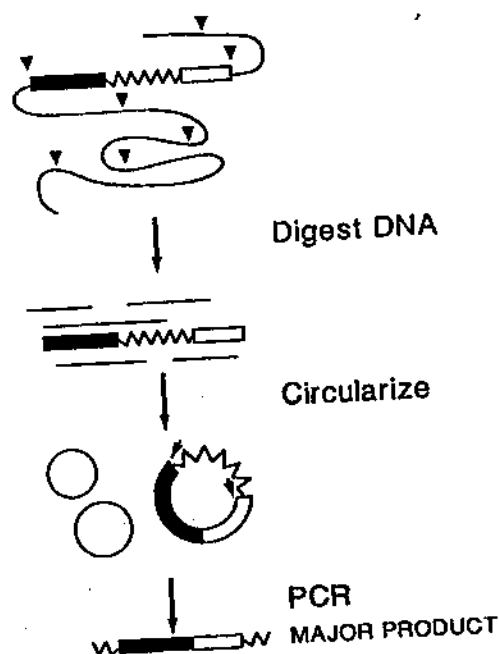
Kỹ thuật PCR neo có một điểm chung là : DNA cloning đi từ một đoạn sequence đã biết rồi, tới đoạn sequence kế cận chưa biết, nhờ sự giúp đỡ của một primer chuyên tính đối với gen tại một đầu dây, và một primer tổng quát tại đầu dây khác. Bởi vì chỉ có primer chuyên tính đối với gen mới có thể neo trong PCR, làm cho dễ dàng hơn việc thu thập một lượng lớn sản phẩm khuếch đại không chuyên tính so với PCR bằng hai primer chuyên tính (Huang và ctv. 1990, Delort và ctv. 1989). Nó khác với PCR chuẩn là chỉ cần một primer đơn cho phản ứng. Trong kỹ thuật PCR neo, một đuôi nhân tạo của dG residues được thêm vào đoạn cuối của DNA mục tiêu. Sự khuếch đại DNA xảy ra khi có primer của một đoạn mã đã biết trước và một primer thứ hai có oligo dC residues (hình 8-1).

8.5.2.2. PCR đảo

Phương pháp PCR đảo có thuận lợi lớn hơn nhờ khuếch đại sequence kế cận chưa biết bằng hai primer chuyên tính đối với gen.

PCR đảo hoạt động theo nguyên tắc tạo dòng dây đơn với đoạn mặt mã DNA. DNA này được tiêu hóa bằng restriction endonuclease tương ứng ở vị trí mà nó tạo ra DNA mục tiêu. Sau đó nó được hoạt động theo chu trình có tính chất monomer. DNA này được khuếch đại lên bằng các primer tương đồng, tại đầu dây của DNA có sequence đã được biết.

Ban đầu là dây 5'-3' của mRNA cho vào phản ứng chuyển mã ngược, thành ssDNA 3'-5'. Người ta tiếp tục tổng hợp dây đơn của cDNA lần thứ hai, cho ra dây đôi dsDNA [5'-3' và 3'-5']. Dây này có chứa một đoạn sequence đã biết rõ. Kỹ thuật cloning làm cho DNA thẳng biến thành cDNA vòng, dây đôi. Kế đến mở vòng, nhờ enzym cắt sequence được biết làm đôi, một nửa ở đầu dây thẳng bên này, một nửa bên kia. Kéo dây thẳng và khuếch đại trong PCR (hình 8-5).



Hình 8.5 : Ứng dụng PCR đảo. Vùng cực trọng được biểu thị bằng đường răng cưa. Các hình hộp được bôi đen và để trắng là những vùng kế cận : trên (upstream) và dưới (downstream), theo thứ tự. Vị trí xác định của những enzyme phân cắt hạn chế được biểu thị bằng hình tam giác. Primer tác động ảnh hưởng annealing tại vùng cực trọng, và sau đó DNA được tổng hợp theo chiều mũi tên (Innis và ctv. 1990)

Vật liệu để tổng hợp dây cDNA đơn lần thứ nhất :

- RNA (theo qui trình phân lập RNA).
- dNTP mix (10mM của mỗi dNTP).
- Các primer ngẫu nhiên, trong nước cất vô trùng $1\mu\text{g} / \mu\text{l}$. Tồn trữ -20°C .

- RNasin.
- Actinomycin D (1mg /ml). Actinomycin D là chất độc, tồn trữ nó trong ống nghiệm được bọc lại bằng giấy nhôm, giữ -20°C.
- MMLV reverse transcriptase.
- 5X first-strand buffer : 250mM Tris-HCl, pH 8.3, 375mM KCl, 50mM MgCl₂, 50mM DTT, và 2,5mM spermidine. Dung dịch này được giữ ở -20°C, có thể kéo dài hơn 6 tháng.

Vật liệu để tổng hợp dây thứ hai :

- 10X second-strand buffer : 400mM Tris-HCl, pH 7.6, 750mM KCl, 30mM MgCl₂, 100mM (NH₄)₂SO₄, 30mM DTT, và 0,5 mg/ml BSA. Dung dịch này được giữ ở -20°C, có thể kéo dài dưới 6 tháng.
- NAD (1mM).
- RNase H (2U/μl).
- *E. coli* DNA polymerase I (5U /μl).
- *E. coli* DNA ligase (1U /μl).
- Nuclease-free H₂O.
- T4 DNA polymerase.
- EDTA (200 mM) pH 8,0.
- GeneClean (Bio 101 Inc., La Jolla,CA).
- TE buffer: 10mM Tris-HCl, pH 7.6, 1mM EDTA. Lọc nước cất.
- DNA standards: 1-mM alipuots của mẫu DNA tinh khiết ở 1, 2.5, 5, 10 và 20 μg/ml trong TE. Tồn trữ -20°C kéo dài hơn 6 tháng.
- TE / ethidium bromide: 2μg/ml. Tồn trữ 4°C / 6 tháng trong tối.

Vật liệu dùng cho tạo cDNA vòng và phân cắt :

- 5X ligation buffer (bổ sung thêm T4 DNA ligase).
- T4 DNA ligase (1U/μl).
- T4 RNA ligase (4μg/μl).
- Hexaminec alt chloride (15 μM).
- GeneClean.

Vật liệu cho PCR đảo :

- PCR regents.
- Deoxyoligonucleotide được tổng hợp.

TÀI LIỆU THAM KHẢO

1. Breslauer KJ, F Ronald, H Blocker, LA Marky. 1986. *Predicting DNA duplex stability from the base sequence*. Proc. Natl. Acad. Sci. USA 83:3746-3750
2. Chou Q, M Russell, DE Birch, J Raymond, W Bloch 1992. *Prevention of pre-PCR mispriming and primer dimerization improves low-copy number amplifications*. Nucleic Acids Res. 20:1717-1723
3. Delort J, JB Dumas, MC Darmon, J Mallet. 1989. *A efficient strategy for cloning 5' extremities of rare transcripts permits isolation of multiple 5'-untranslated regions of rat tryptophan hydroxylase mRNA*. Nucleic Acids Res. 17:6439-6488
4. Drillon PJ, CA Rosen. 1993. *Use of polymerase chain reaction of the rapid construction of synthetic genes*. Methods in Mol. Biol. 15:263-268
5. Dveksler GS, CW Dieffenbach, CB Cardellicchio, K McCuaig, MN Pensiero, GS Jiang, N Beauchemin, KV Holmes. 1993. *Several members of the mouse carcinoembryonic antigene-related glycoprotein family are functional receptors for the coronavirus mouse hepatitis virus-A59*. J. Virol. 67:1-8
6. Freier SM, R Kierzek, JA. Comeger, N Sugimoto, MH Caruthers, T Neilson, DH Turner. 1986. *Improved free-energy parameters for predictions of RNA duplex stability*. Proc. Natl. Acad. Sci. 83:9373-9377
7. Huang SH, CH Wu, B Cai, J Holcenberg. 1993. *cDNA cloning by inverse polymerase chain reaction*. Methods in Mol. Biol. 15:349-356
8. Huang SH, YY Hu, CH Wu, J Holcenberg. 1990. *A simple method for direct cloning cDNA sequence that flanks a region of known sequence from total RNA by applying the inverse polymerase chain reaction*. Nucleic Acids Res. 18:1922
9. Keohavong P, WG Thilly. 1989. Proc. Natl. Acad. Sci. USA 86:9253
10. Kwok S, DE Kellog, N McKinney, D Spasic, L Goda, C Levenson, JJ Sninsky. 1990. *Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus 1 model studies*. Nucleic Acids Res. 18:999-1005
11. Liang P, A Pardee. 1992. *Differential display of eukaryotic mRNAs by PCR*. Science 257:967-971
12. Mack DH, JJ Sninsky. 1988. *A sensitive method for the identification of uncharacterized viruses related to known virus groups: Hepadnavirus model system*. Proc. Natl. Acad. Sci. 85:6977-6981

13. McConaughty BL, CL Laird, BJ McCathy. 1969. *Nucleic acid reassociation in formamide*. Biochemistry 8:3289-3295
14. Mullis K, F Faloona, S Scharf, R Saiki, G Horn, H Erlich. 1986. *Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction*. Cold Spring Harbor Symp. Quant. Biol. 51:263-273
15. Mullis KB, FA Faloona. 1987. *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction*. Methods Enzymol. 155:335-350
16. Oste C. 1989. *Amplifications* 1:10
17. Ou CY, S Kwok, SW Michell, DH Mack, JJ Sninsky, JW Krebs, P Feorino, D Warfield, G Schochelman. 1988. *DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells*. Science 239:295-297
18. Ruano G, DE Brash, KK Kidd. 1991. *Amplifications* 7:1-4
19. Rychlik W, WJ Spencer, RE Rhoads. 1990. *Optimization of annealing temperature for DNA amplification in vitro*. Nucleic Acids Res. 18:6409-6412
20. Saiki RK, SJ Scharf, F Faloona, KB Mullis, HA Ehrlich, GT Horn, N Arnheim. 1985. *Enzymatic amplification of β -gl in genomic sequences and restriction site analysis for diagnosis of sickle cell anemia*. Science 230:1350-1354
21. Suggs SV, T Hirose, EH Myake, MJ Kawashima, KI Johnson, RB Wallace. 1981. *Using purified genes*. ICN-UCLA Symp. Mol. Cell Biol. 23:683-693
22. Verma IM, GF Temple, H Fan, D Baltimore. 1972. *In vitro synthesis of double-strand DNA complementary to rabbit reticulocyte 10S RNA*. Nature 235:163-169
23. William JGK. 1990. *DNA polymorphism amplified by arbitrary primers are useful as genetic markers*. Nucleic Acids Res. 18:6531-6535
24. Wu DY, W Ugozzoli, BK Pal, J Qian, RB Wallace. 1991. *The effect of temperature and oligonucleotide primer length on specificity and efficiency of amplification by the polymerase chain reaction*. DNA Cell Biol. 10:233-238

CHUỖI MÃ DI TRUYỀN DNA [DNA SEQUENCE]

9.1. NGUYÊN TẮC CƠ BẢN CỦA DNA SEQUENCING

Kỹ thuật DNA sequencing là kỹ thuật xác định tất cả những hợp phần do nucleotide hình thành nên phân tử DNA chuyên tính nào đó (Alphey 1997). Khả năng đọc được chuỗi mã DNA gắn liền với nhiệm vụ trung tâm của công nghệ sinh học phát triển. Những kỹ thuật sequencing cũng chỉ mới được hoàn thiện trong những năm gần đây. Các công trình đầu tiên được công bố của Maxam và Gilbert (1977), Sanger và ctv. (1977). Do đó phương pháp sequencing có tên phương pháp Sanger, hoặc Maxam và Gilbert. Nhưng trong hơn 20 năm qua, kỹ thuật này được cải tiến rất nhiều. Đặc biệt với sự trợ giúp của computer, kỹ thuật huỳnh quang, tiến bộ của PCR, và kỹ thuật điện di, phương pháp sequencing trở thành công cụ rất có giá trị, làm nền tảng cho phân tích genome.

Cấu trúc của phân tử DNA là một dây đôi xoắn, do hai dây đơn hợp lại. Trên mỗi dây, có một chuỗi các base, tập hợp của bốn nucleotic acid : A, C, G, T. Mỗi base được gắn với đường deoxyribose và liên kết với nhau thông qua gốc phosphate. Do đó đơn vị cơ bản của DNA là base, đường và phosphate, được hiểu với thuật ngữ nucleotide.

Phương pháp sequencing RNA được phát triển sớm hơn DNA, nhưng hiện nay RNA ít được sequenced một cách trực tiếp. Trong tổng hợp cDNA, người ta có thể đọc sequence của cDNA, rồi suy ngược lại RNA gốc của nó.

DNA bao gồm bốn base khác nhau, liên kết với nhau trên đường thẳng. Chúng ta có thể biểu hiện chúng theo mô thức sau: 5'-ACGT-3'.

DNA sequence được viết gọn là 5'-3' theo phương pháp cổ điển, và hiện nay có thể viết đơn giản hơn ACGT.

Theo tính chất bắt cặp của hai dây đối nhau, A với T, C với G, nên chúng ta có thể biểu hiện chúng ở dạng dây đôi: 5'-ACGT-3'

3'-TGCA-5'

Giả dụ cho A = 0, C = 1, G = 2, T = 3

5' A C G T 3'

0 1 2 3

Phép toán nhị nguyên sẽ viết lại là:

00 01 10 11

Như vậy biểu thị cho dây đôi sẽ là 00 01 10 11

11 10 01 00

Đó là nguyên tắc cơ bản để thiết lập chương trình computer trong kỹ thuật sequencing.

Phương pháp nhanh nhất để có được một bản đồ gen chi tiết là ứng dụng computer trong khi đọc mã. Sự xác định với computer về một vùng mã của protein nào đó [được gọi bằng thuật ngữ: open reading frame hoặc ORF] giống như phương pháp tìm DNA và tìm dữ liệu protein. Điều này rất quan trọng trong việc tìm hiểu chức năng và cấu trúc của gen, cũng như các clone của nó. Hơn nữa, DNA sequence cần cho những phân tích chi tiết về vùng không mang mã 3' và 5' của một gen nào đó. Thông tin về DNA sequence vô cùng cần thiết cho sự phát sinh thể đột biến [mutagenesis].

Số lượng nhỏ của thông tin DNA sequence, còn gọi là sequence tag site, hoặc STS, hay expressed sequence tag, hoặc EST, là phần căn bản của phương pháp lập bản đồ và điều khiển các đoạn DNA đại phân tử thành YAC [nhiễm thể nhân tạo làm bằng men] hay những cosmid.

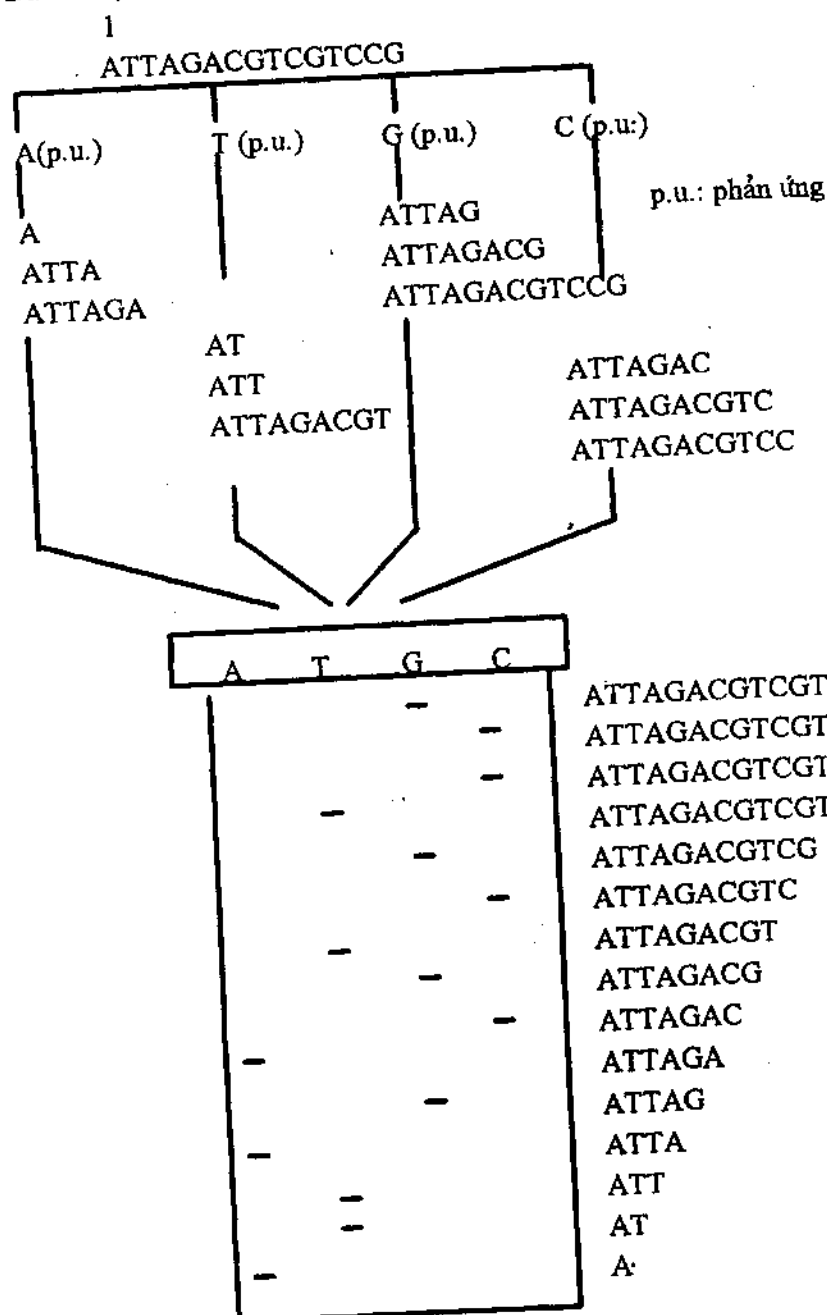
Kỹ thuật DNA sequencing dựa trên kỹ thuật điện di, sử dụng gel loại polyacrylamide có độ phân giải cao. Người ta gọi đó là "sequencing gels", có khả năng chạy các oligonucleotide dây đơn, với độ dài 500 bp, nó có thể phân biệt được đến mức deoxynucleotide. Trong thực hành, đối với một vùng đang được đọc mã nào đó, một loạt các oligonucleotide được đánh dấu, cố định một đầu, phân biệt về kích thước với đầu khác bằng mỗi deoxynucleotide đơn. Chìa khóa để xác định chuỗi mã di truyền là tất cả các oligonucleotides được kết thúc bằng một trong bốn phân tử A, C, G hoặc T. DNA sequence có thể được đọc trực tiếp từ bốn ladder này của oligonucleotide.

Thuật ngữ sequence được sử dụng có tính chất quốc tế hóa, nên có thể không cần dịch ra tiếng Việt, để tiện việc diễn giải.

9.1.1. Phương pháp hóa học (Maxam và Gilbert)

Phương pháp hóa học dựa trên cơ sở phân cắt hóa học tại vị trí đặc biệt của base, tạo ra một phân tử DNA có đầu được đánh dấu, hình thành một loạt phân tử đánh dấu tận cùng bằng một base (thí dụ A, hoặc C, hoặc G, hoặc T). Sequence của DNA có thể được đọc từ kết quả của ladder (thang chuẩn) còn gọi là "sequence ladder". Phương pháp này còn được gọi là phương pháp kết thúc theo chuỗi, xác định base nào đó từng điểm cuối cùng trong chuỗi mã, cứ như vậy cho 3 base còn lại (hình 9-1).

Đầu cố định



Hình 9-1: Nguyên tắc DNA sequencing: một dây đơn oligonucleotide có đánh dấu phóng xạ, được chia thành bốn phản ứng. Trong mỗi bốn phản ứng, oligonucleotide có một đầu cố định và một đầu có phân tử A, T, G và C. Sản phẩm của mỗi phản ứng được điện di trên polyacrylamide có độ phân giải cao. Chụp trên gel hình ảnh của DNA sequence theo phương pháp autoradiography.

Bảng 9.1 : Phản ứng chuyên tính của base trong phương pháp hóa học

Hóa chất	Base bị ảnh hưởng
Dimethyl sulfate pH	8,0 G
Piperidine formate pH 2,0	A + G
Hydrazine	C + T
Hydrazine + 1,5 M NaCl	C
Nhiệt độ nóng (90oC) 1,2 M NaOH	A > C

Những hóa chất này phản ứng với DNA để thay đổi base chuyên tính của nó. Piperidine nóng (1M, nước 90°C) chỉ cắt phosphate-đường của DNA, ở đó có một base đã được thay đổi. NaOH nóng (A > C) cho một sự phân cắt mạnh tại A và yếu hơn tại C. Phản ứng bổ sung cho từng base chuyên tính đều có thể xảy ra.

Ưu điểm của phương pháp :

- Sequence có thể được xác định trên cơ sở bản đồ restriction.
- Sequence có thể được xác định rất gần vị trí đánh dấu (trong khoảng 2-3 base).
- Sequence nhận được từ phân tử DNA nhiều hơn là từ copy của enzym.

Khuyết điểm :

- Sequence thường được nhận rất ít.
- Các phản ứng xảy ra chậm và độ tin cậy thấp.
- Phản ứng đòi hỏi nhiều hóa chất có tính chất ngẫu nhiên.

9.1.2. Phương pháp đọc chuỗi mã Sanger

Phương pháp sequencing do Sanger và cộng sự viên tìm ra năm 1977. Họ dùng một DNA polymerase để tổng hợp một bản sao có tính chất bổ sung, từ một dây template của DNA. Phương pháp này còn được gọi là phương pháp Sanger dideoxy.

Các vật liệu cơ bản bao gồm: primer

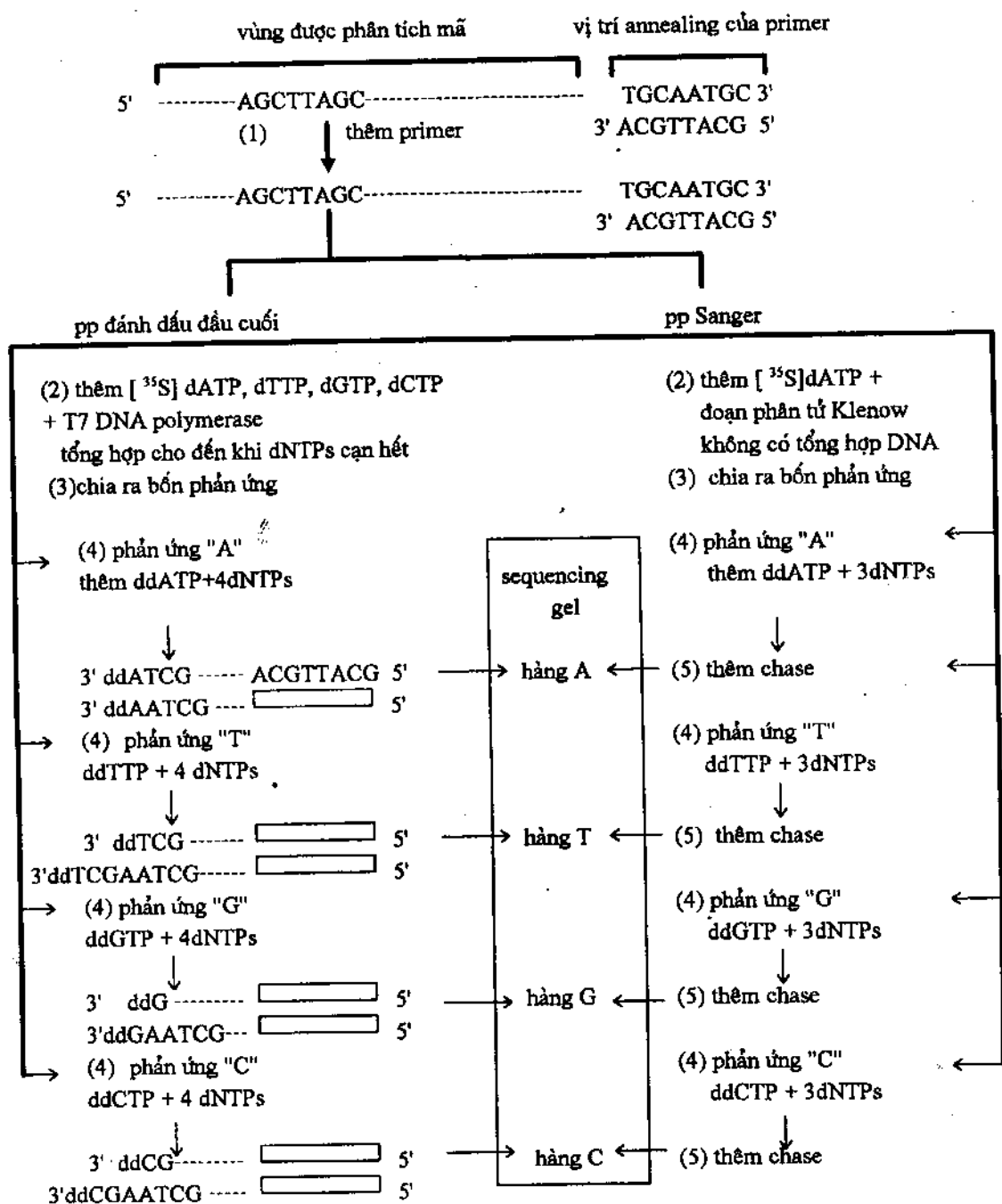
DNA polymerase

Label

dNTP và ddNTP

dây template DNA

hệ thống điện di



Hình 9-2: Phương pháp dideoxy sequencing

Mỗi dây đơn của DNA được tác động với một primer tương ứng trong phản ứng polymer. Cho thêm vào đoạn phân tử Klenow và dATP được đánh dấu phóng xạ (³⁵S). Phản ứng được chia nhỏ thành bốn giai đoạn, và 3 dNTP còn lại cũng như ddATP, ddCTP, ddGTP, ddTTP được thêm vào. Sự tổng hợp DNA xảy ra, kết thúc bằng sự gắn vào của

ddNTP. Một chất săn đuổi (chase) trong step 5, theo bốn dNTP làm kéo dài chuỗi chưa được kết thúc bằng ddNMP của phân tử DNA có trọng lượng phân tử cao (hình 9-2).

Ở đây, DNA polymerase được dùng để tổng hợp một copy bổ sung của dây đơn DNA. Polymerase không thể khởi động chuỗi mã DNA. Sự kéo dài chuỗi xảy ra ở đầu 3' của primer, và nó thực hiện annealing trên dây DNA. Sự phát triển của chuỗi bao gồm sự hình thành cầu nối phosphodiester giữa gốc hydroxyl 3', tại đầu dây của primer và tại vị trí gốc phosphate 5' của deoxynucleotide. Do đó, sự phát triển của chuỗi theo hướng 5'-3'.

Phương pháp dideoxy khai thác khả năng của DNA polymerase nhằm sử dụng 2', 3'-dideoxynucleoside triphosphate (ddNTPs) làm cơ chất. Khi ddNMP được gắn vào tại đầu 3' của chuỗi mã primer, thì sự kéo dài chuỗi mã được kết thúc bằng G, A, T hoặc C bởi vì chuỗi primer bây giờ thiếu một gốc hydroxyl 3'.

Để tạo ra bốn "ladder" trong sequencing, chỉ có một trong bốn ddNTP được hiện diện trong một của 4 phản ứng. Tỷ lệ ddNTP : dNTP của mỗi phản ứng được điều chỉnh sao cho phần của chuỗi primer kéo dài ra sẽ kết thúc tại mỗi base trên dây template, tương ứng với ddNTP được thêm vào. Theo kiểu này, mỗi một trong bốn phản ứng kéo dài chuỗi đều chứa đựng quần thể của chuỗi primer được kéo dài. Tất cả có một đầu cố định 5', được xác định bởi primer quần lấy nó và một đầu 3' tự do, kết thúc tại một dideoxynucleotide đặc biệt nào đó của nó (Current Protocol 1993).

Phương pháp dideoxy yêu cầu sequencing trên dây đơn template, để primer quần lấy nó. Dây đơn này rất dễ tạo ra bằng cách sử dụng vectơ M13, một phage của *E. coli*. Phương pháp dideoxy có thể thực hiện trên dây đôi DNA, nếu lần biến chất (denature) đầu tiên được thực hiện với alkali và nhiệt độ nóng (Hattori và Sasaki 1986, Chen và Seeburg 1985).

$[\alpha^{35}\text{S}]$ dATP đánh dấu bằng đồng vị phóng xạ, hiện được dùng nhiều hơn $[\alpha^{32}\text{P}]$ dATP vì các lý do sau đây :

- Khả năng tải năng lượng thấp của $\alpha^{35}\text{S}$ tạo ra kết quả bằng chụp trên phim sắc nét hơn so với ^{32}P , giúp chúng ta đọc được nhiều sequence hơn, ở các vùng trên của gel.
- Khả năng tải năng lượng thấp của $\alpha^{35}\text{S}$ ít tạo ra hiện tượng gãy đứt gốc đường-phosphate của phân tử DNA.
- Trong thực tế, sản phẩm của $\alpha^{35}\text{S}$ có thể tồn trữ ở -20°C trong nhiều tuần, trong khi sản phẩm của ^{32}P phải được điện di trong ngày hôm đó.
- Người sử dụng có thể chấp nhận liều lượng thấp của ^{35}S hơn ^{32}P .

Tuy nhiên, ^{32}P có ưu điểm về thời gian thể hiện ngắn, và đặc biệt trong vài trường hợp, nó rất hữu dụng, như kiểm tra sự hình thành plasmid, trong khi các giải pháp về sequencing

không có kết quả. Gần đây, nhờ những cải tiến mới, [$\alpha^{32}\text{P}$]dATP trong phản ứng sequencing đã tối đa hóa năng lượng tải cao hơn 50% so với $\alpha^{35}\text{S}$ (Zagursky và ctv. 1991)

9.2. CHUẨN BỊ DÂY TEMPLATE

Giai đoạn đầu tiên của kỹ thuật sequencing là chuẩn bị dây đơn DNA, dây template đối xứng với dây gốc, nơi phản ứng quán (annealing) của primer sẽ xảy ra.

Cách đơn giản nhất là tạo các dòng phụ (subclone) của đoạn phân tử DNA trong vectơ M13, tinh khiết phage và ly trích dây đơn DNA. Tuy nhiên, vectơ M13 có nhược điểm là nó có thể sắp xếp lại, hoặc xóa bỏ hẳn những phần gắn vào có phân tử lớn, và nó có thể gặp nhiều khó khăn trong quá trình tinh khiết thể lặp lại của dây đôi (replicative form), cần cho restriction enzyme, và kỹ thuật subcloning.

Hiện nay, người ta thường sử dụng những vectơ có thuật ngữ là "phagemid". Đó là những plasmid mang hai vùng "ori" (nguồn gốc của tự tái bản): một của plasmid và một của phage. Sự có mặt của phage ở đây có tên là "helper phage", với dạng DNA dây đơn và các thành phần khác của phage, như M13 K07, mã hóa protein cần cho quá trình tự tái bản của phage, cũng như sự phát triển của phage, nhưng tự thân nó sẽ phát triển rất kém. Helper phage giúp cho sự tạp nhiễm của dây DNA giảm đi rất nhiều. Ngoài ra, nó không có vùng bổ sung cho các primer tổng quát, nên nó không ảnh hưởng đến phản ứng sequencing. Phagemid là vectơ có vị trí MCS (viết tắt từ chữ multiple cloning site) cho phép sự cloning đa dạng hơn, nằm bên cạnh những chuỗi mã của primer có tính chất tổng quát (universal primers). Đặc điểm chính của vectơ này là một ori của plasmid và một ori của phage có dây đơn, một marker có tính chọn lọc (thí dụ tính kháng ampicillin), và một loạt các vị trí phân cắt giới hạn cho cloning (MCS). Vùng nguồn gốc tự tái bản của phage cho phép một dây đơn được thuần khiết, theo sau đó là sự hoạt động của helper phage.

9.2.1. Phương pháp sử dụng phagemid

- Chuyển nạp phagemid DNA vào một nòi *E. coli* tương ứng.
- Chủng 1ml LB có chứa 50-100g/ml ampicillin với một khuẩn lạc, cho phát triển qua đêm ở 37°C, có lắc mẫu.
- Trộn lại trong cùng một ống nghiệm vô trùng 10-15ml:

LB có 50-100g / ml ampicillin	2,5 ml
Môi trường nuôi cấy vi khuẩn có phagemid	50 μ l
Helper phage " M13 K07 "	3 μ l
- Lắc mẫu bằng ly tâm 250-300 vòng/phút ở 37°C, trong 1 giờ.
- Thêm kanamycin cho đến 100g/ml.
- Nuôi cấy trong điều kiện lắc ở 37°C, trong vòng 6-8 giờ hoặc qua đêm.

9.2.2. Phương pháp sử dụng M13

- Phát triển 1ml môi trường nuôi cấy trên các tế bào chủ tương ứng đến bão hòa.
- Trộn chung trong cùng một ống nghiệm vô trùng 10-15ml : LB (2,5ml) và môi trường nuôi cấy vi khuẩn (50 μ l).
- Đung vào que vô trùng, hoặc vòng kim loại trên plaque của M13, rồi dùng nó để đưa M13 vào môi trường.
- Môi trường nuôi cấy được lắc ở 37°C, trong 5-7 giờ.

Hai qui trình này được sử dụng cho phương pháp dideoxy. Qui trình cho sản phẩm PCR sẽ được trình bày ở các phần sau (sequencing trực tiếp các sản phẩm PCR). Phần này sẽ có những ứng dụng nhiều hơn.

Ngoài ra, chúng ta còn có các qui trình chuẩn bị template của plasmid, của lambda, cosmid và P1, hoặc kỹ thuật sequencing có tính chất nửa tự động (semi-automated).

9.3. ĐIỆN DI TRÊN GEL

Những DNA polymerase được dùng trong phản ứng sequencing có thể tổng hợp thành một loạt các phân tử với trên 1000 base từ primer, cho chúng ta một dây template' khá tốt. Điều này có nghĩa là hơn 1000 base của thông tin di truyền có đầy đủ khả năng cho một phản ứng sequencing. Trong thực tế, điện di trên gel tiếp sau đó không thể nào giải quyết nổi toàn bộ các phân tử, làm hạn chế khối lượng thông tin di truyền có thể nhận được. Số lượng thông tin di truyền có thể được đọc hết và dễ dàng thu nhận. Điều này tùy thuộc vào hệ thống gel được sử dụng, và phương pháp điện di cải tiến.

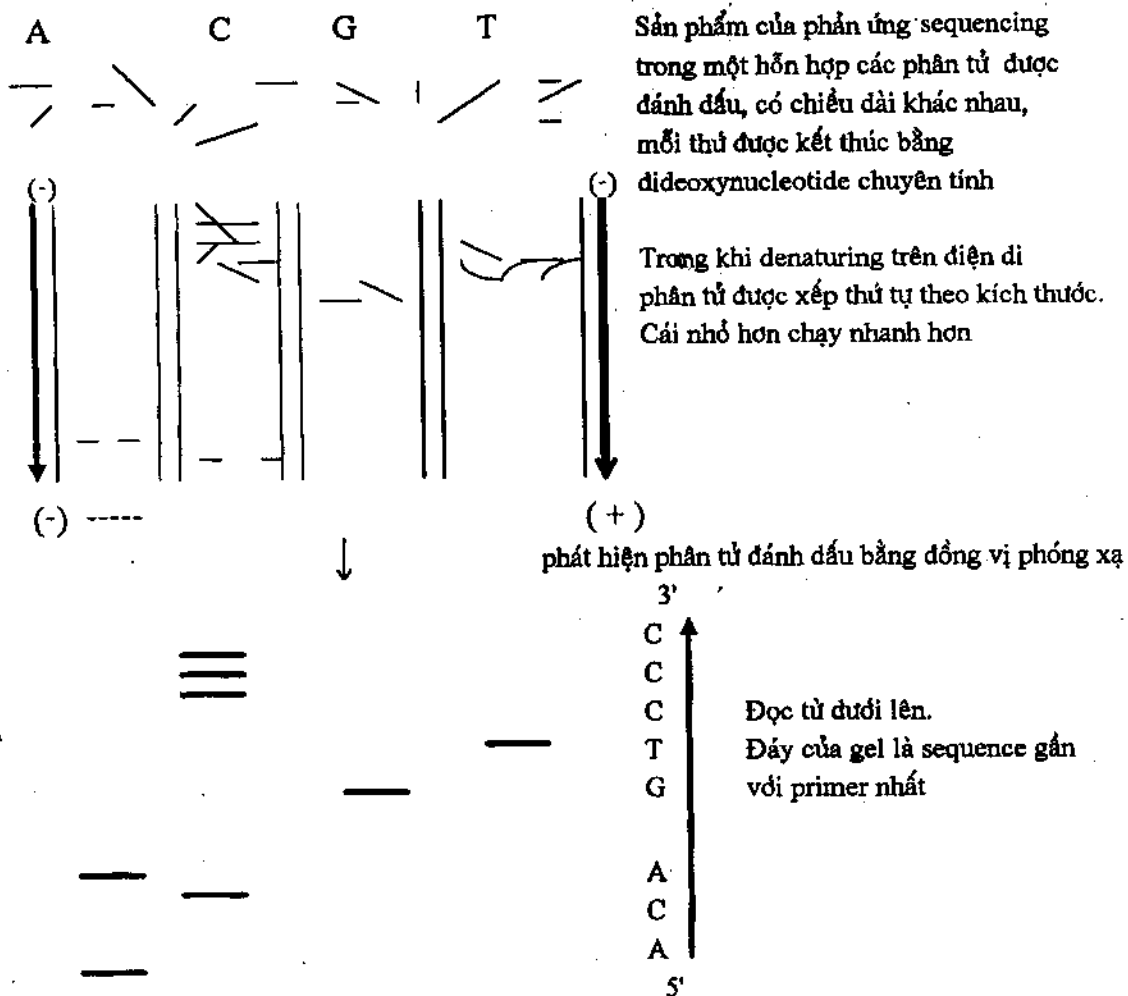
Hệ thống gel cho kỹ thuật sequencing sử dụng nguồn điện có hiệu thế cao (1500-3000 V). Do đó nguyên tắc an toàn trong phòng thí nghiệm luôn luôn được đặt ra. Ngày nay, với các máy sequencer rất đắt tiền, hiện đại, nguyên tắc này được cải tiến đi nhiều lần.

Tấm gel được làm bằng thủy tinh (hai tấm thủy tinh có mặt phẳng đúng tiêu chuẩn), kẹp giữa polyacrylamide gel. Tấm thủy tinh cần được rửa cẩn thận và làm khô trước khi dùng để chạy gel.

Lược (comb) có răng nhọn, nên có tên là lược răng cá mập, tạo ra các giếng (well) đúng yêu cầu để loading tốt hơn lược thông thường, cho nhiều sequence hơn đọc được trên băng.

Gel có bề dày 0,4mm và rộng 30-40 cm. Chiều dài tiêu chuẩn của gel là 40-50cm. Gel dài hơn 100cm khi chạy điện di phải có thêm công cụ hỗ trợ (như S1, Life Tech.), nhưng khi loading, cố định, sấy khô, và thể hiện mẫu sẽ rất khó khăn.

Nhiệt độ để gel hoạt động tốt là 40 - 50°C.



Hình 9.3: Sequencing gels

9.4. PHƯƠNG PHÁP KHÔNG DÙNG CHẤT PHÓNG XẠ

Do tính chất nguy hiểm đến sức khỏe, người ta cải tiến không ngừng tìm ra phương pháp không dùng phóng xạ để đánh dấu. Phương pháp có tên là "Promega's Silver Sequence" sử dụng bạc (Ag) để nhuộm mẫu, rồi sau đó chụp bằng phim đặc biệt (Kodak EDF).

Ưu điểm của phương pháp này là :

- Nó không cần chất phóng xạ.
- Thời gian chạy trên gel để có thông tin di truyền thường được rút ngắn.

Nhược điểm :

- Số liệu thu thập được không tốt bằng phương pháp đánh dấu phóng xạ.
- Mất nhiều thời gian để chuẩn bị.

Phương pháp không sử dụng chất phóng xạ được áp dụng trên máy sequencer bán tự động (semi-automated) với hệ thống GATC 1500 (GATC GmbH). Hệ thống này chuyển tải một cách tự động các sản phẩm trong phản ứng điện di vào một màng nylon, bằng cách xê dịch màng dần xuống vùng đáy của gel. Theo cách đó, mỗi phân tử phải di chuyển suốt chiều dài của gel, tạo thuận lợi tối đa cho việc giám định. Phương pháp mới này cho phép nhiều thông tin di truyền hơn trong một "giếng" [trên 1000 base], nếu mọi điều kiện được thỏa mãn. Hệ thống sequencing này rất đắt tiền.

Chúng ta sẽ thảo luận dưới đây các hệ thống bán tự động, ABI 377, LI-COR

9.4.1. Máy sequencer bán tự động

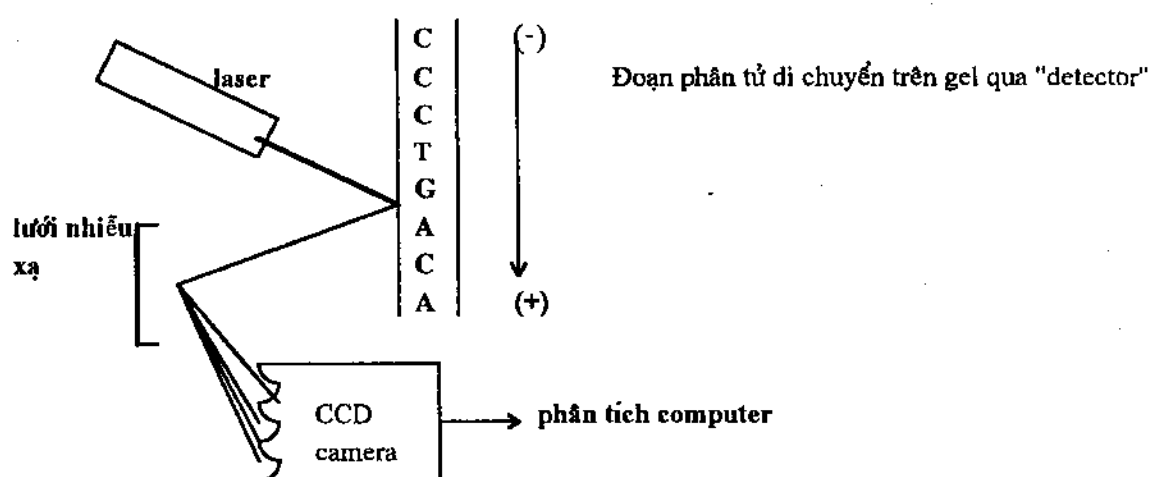
Thực ra, người ta có thể gọi công cụ này là tự động (automated) hoặc bán tự động. Trong phương pháp thủ công có tính chất tiêu chuẩn hóa, người ta phải dùng chất đồng vị phóng xạ để đánh dấu, và quan sát băng (band), bằng phương tiện autoradiography (máy chụp phóng xạ), sử dụng các photodetector cực nhạy, để phát hiện các đoạn phân tử có đánh dấu với phẩm nhuộm huỳnh quang, được quét trên laser, gần đáy của gel (Ansorge và ctv. 1987, Middendorf và ctv. 1992). Phương pháp tự động hoặc bán tự động sử dụng các số liệu lưu trữ có trong phần mềm phân tích [computer], tự vận hành toàn bộ quá trình, từ vị trí loading mẫu đến suốt quá trình sequencing trên gel.

Hiện nay, hầu hết các sequence mới đều được xác định bằng máy này. Nó không những có tính chất dễ thực hành trong phòng thí nghiệm mà còn tiết kiệm nhiều thời gian khởi động trong kỹ thuật sequencing để phân tích genome. Gần đây, người ta đã xác định toàn bộ sequence của men *Saccharomyces cerevisiae* vào tháng 5/1996 (Alphey 1997), và dự kiến trong vài năm nữa, người ta sẽ hoàn thành việc xác định sequence của một số sinh vật có giá trị trong nghiên cứu khoa học, nông nghiệp, y khoa, bao gồm cả genome của người (*Homo sapiens*). Giá cả của máy sequencer biến thiên từ 60.000 đến 120.000 US đô la.

9.4.2. ABI 377

Máy ABI 377 có thể phát hiện cùng một lúc hiện tượng huỳnh quang ở bốn độ dài sóng khác nhau, tạo một kết hợp cho phép truyền đi bốn màu huỳnh quang khác nhau. Sự phát hiện cùng một lúc như vậy được hoàn thiện nhờ phân biệt rõ ánh sáng dẫn truyền với một lưới nhiễu xạ, cho phép việc phát hiện này với nhiều độ dài sóng, nhờ dụng cụ tinh vi để chụp. Đó là camera có tên kỹ thuật CCD, viết tắt từ chữ "charge-couple device". Mỗi phẩm màu tương ứng với một trong bốn dideoxynucleotide, để đánh dấu. Các phản ứng chạy trên một hàng đơn, sao cho màu của mỗi băng đều chạy qua detector (hình 9-4).

Máy ABI 377 chiếu sáng gel bằng một argon laser. Máy điều khiển và theo dõi huỳnh quang từ các DNA đã được đánh dấu. Sự thu thập số liệu thực như vậy có nghĩa là không có cơ hội để điều chỉnh tín hiệu biến đổi - khác với phương pháp sequencing bằng tay, ở đó chúng ta có thể dừng lâu hơn trên máy chụp bức xạ, nếu cần.



Hình 9-4 : Thu thập số liệu thực trên máy ABI 377

Những dấu hiệu mờ nhạt, không rõ cần có detector có tính nhạy cảm cao hơn, làm cho hệ thống trở nên cực nhạy. Có hiện tượng nhiễu tạp do fluorochrome ngoại sinh. Những thể tạp như vậy, có thể có từ nguồn không độc hại như giấy thấm, giấy lọc.

Quan sát lần đầu, số liệu thô của ABI 377 không có vẻ hấp dẫn lắm, nhưng sau đó là một kết quả đầy sức thuyết phục. Mỗi hàng là ladder của những băng nhuộm màu, mỗi màu đại diện cho một nucleotide.

Có hai hóa chất trong ABI 377, đó là primer nhuộm màu (dye primer) và terminator nhuộm màu. Đặc tính hóa học của dye terminator là tạo ra sự biến thiên lớn hơn tại đỉnh đường parabol, so với dye primer. Đỉnh thấp của G thường tạo ra các sự cố, đặc biệt khi nó theo sau là A. Mỗi phương pháp sử dụng một trong hai chất này đều có những ưu và nhược điểm riêng.

Dye terminator : mỗi một trong bốn dideoxynucleotide được đánh dấu với một màu fluorochrome khác nhau. Dye terminator có thể được dùng trong chu kỳ sequencing hoặc trong một lần sequencing (one-pass sequencing). Ưu điểm chính của phương pháp này là chỉ cần dây template có trọng lượng phân tử thấp. Điều này quan trọng cho sequencing sản phẩm PCR. Chúng ta có thể dùng công cụ thử nghiệm nhanh (kit) ABI PRISM Ready Reaction kits, Perkin-Elmer, để hoàn thiện phản ứng sequencing.

Dye primer: màu nhuộm huỳnh quang gắn dính trong primer (Smith và ctv. 1986). Điều này có nghĩa là, một phản ứng sequencing thông thường, cần bốn phản ứng khác nhau, một với một primer đã được đánh dấu bằng màu huỳnh quang. Tổng hợp một primer có đánh dấu huỳnh quang đắt tiền hơn một primer không đánh dấu. Do đó một bộ primer cho kỹ thuật dye primer sequencing sẽ đắt gấp bốn lần phương pháp dye terminator sequencing. Phản ứng sequencing phải được thực hiện trên bốn ống nghiệm khác nhau, rồi gộp chung lại trên gel khi điện di, so với phương pháp trên chỉ cần một ống nghiệm, ít tốn

công khởi động hơn. Nhưng phương pháp dye primer có ưu điểm là tín hiệu ghi nhận được rất đồng nhất, từ một base này sang một base khác. Dye primer sử dụng AmpliTaq DNA polymerase, FS, hoặc sequenase rất thành công, trong khi phương pháp dye terminator thì không.

9.4.3. LI-COR

Giống như ABI, máy LI-COR là một sequencer tự động, phát hiện phân tử DNA được đánh dấu bằng huỳnh quang, nhờ quét chùm tia laser gần đáy của gel.

Máy LI-COR sử dụng kính hiển vi huỳnh quang để tìm thể đánh dấu bằng huỳnh quang (Middendorf và ctv. 1992). Điều này cho phép chúng ta phát hiện tập trung trên một hàng của gel. Sự tập trung này có thể được kiểm soát một cách tự động, trên cơ sở màu huỳnh quang khác nhau của gel tương ứng với các tấm kính thủy tinh. Chỉ có một độ dài sóng được khám phá, cho nên số liệu có thể được thu thập nhanh hơn trong mỗi lần quét laser.

Primer được đánh dấu theo hệ thống labeling tiêu chuẩn, dùng dATP đánh dấu với phẩm màu của tia hồng ngoại. Nó tương đương với hệ thống đánh dấu bằng [-35S]dATPS, và nó cho phép hiểu rõ hơn qui trình sequencing bằng tay hoặc sequencing theo chu kỳ, với những cải tiến nhỏ.

Hệ thống LI-COR 4000LS sử dụng tấm kính 66cm làm chuẩn, trong khi ABI 377 sử dụng tấm kính 36 hoặc 48 cm.

LI-COR 4000LS có thể cho hơn 800 base trên một mẫu ở độ chính xác 99%, nhưng chỉ có thể chạy 7 đến 11 mẫu trên một gel, trong khi ABI 377 có thể chạy 36 đến 64 mẫu / gel. Nhưng LI-COR thường có giá rẻ hơn một chút.

9.5. PHÂN TÍCH COMPUTER CHUỖI MÃ DNA VÀ PROTEIN

Khả năng để xác định chuỗi mã DNA ngày càng trở nên phổ biến tại các phòng thí nghiệm sinh học phân tử, bởi sự hoàn thiện không ngừng về phương pháp và dụng cụ sao cho đơn giản và chính xác hơn. Đặc biệt là những thành tựu trong lĩnh vực thiết kế phần mềm của máy tính. Database của GenBank DNA sequence hiện có hơn 90 triệu nucleotide, đại diện cho hơn 70.000 loci. Trong vòng 4 năm qua, số nucleotid được báo cáo gia tăng trung bình 50% mỗi năm (Current Protocol 1992). Chỉ riêng năm 1991, GenBank đã phát triển hơn 26 triệu nucleotide. Tốc độ phát triển của "database" này đi trước tốc độ xác định sequence của DNA, bởi vì nó có phương tiện có tính chất tự động hóa nhiều hơn và rẻ tiền hơn. Giá của hệ thống computer thường phù hợp với kế hoạch đầu tư cho một la bô, với xu thế ngày càng rẻ. Điều này cho phép nhiều nhà nghiên cứu tham gia vào việc sưu tập, đánh giá database điện tử này, và phân tích sequence trên computer. Với sự phát triển của internet trên toàn cầu, người ta có thể nói rằng kỷ nguyên thông tin mới về sinh học phân tử

phát huy tác dụng rất hiệu quả.

Trước khi phân tích một sequence nào đó của DNA, thông tin phải ở trạng thái mà puter có thể hiểu được. Thông thường, thông tin về chuỗi mã DNA nằm trong một tập tin (file) của computer. Trước hết, người ta mô tả trình tự để nhập số liệu của chuỗi mã ở trang đầu tiên. Sau đó, người ta giới thiệu các mục: phương pháp phân tích và điều chỉnh sequence, xây dựng bản đồ restriction, thiết kế oligonucleotide, xác định vùng mang tín hiệu protein, tìm kiếm thể tương đồng (homology). Phần cuối, người ta mô tả phương pháp xử lý thông tin từ sequence database và những nguồn thông tin khác.

9.5.1. Nhập số liệu

Đối với những chương trình có chuỗi mã ngắn, việc nhập số liệu tương đối đơn giản, cũng như xử lý số liệu trên word processor hoặc text editor của Macintosh hay IBM đều có thể được. Khi kích thước của DNA tăng dần lên, người ta phải dùng hệ editor đặc biệt, để nhập số liệu.

Word processor : một chuỗi mã DNA có thể được nhập vào computer bằng cách sử dụng word processor hay text editor, bằng cách tạo ra document mới, rồi đánh chữ các số liệu của sequence. Những document này được lưu trong text hoặc ASCII. Bởi vì chương trình phân tích sequence không thể dịch ra trên file bằng word processor.

Sequence editor : tất cả các chương trình phân tích đều có chứa một vài dạng của sequence editor. Đây là công cụ có hiệu quả nhất để nhập số liệu của chuỗi mã DNA, bởi vì nó cung cấp một hồ sơ về sequence theo một format chính xác, có bình luận khác với sequence mà nó đang phân tích. Thí dụ chương trình DNA Strider chạy trên Macintosh, sử dụng sequence editor. Nó còn vẽ được bản đồ restriction, mở reading frame, và giải mã DNA sequence thành amino acid sequence.

9.5.2. Máy đọc trên gel tự động

Máy gồm có một scanner và một computer. Máy đọc trên gel tự động với một scanner có độ phân giải cao, hoặc một máy video camera điện tử để chụp biểu đồ bức xạ của một sequencing gel, mà gel này đã được chuẩn bị bằng phương pháp thông thường. Hình ảnh được mã hóa bằng ký hiệu điện tử trên phim, rồi được computer phân tích. Người ta thiết kế phần mềm này nhằm mục đích xác định vị trí của hàng (lane) và vị trí của băng (band) ở trong hàng. Máy đọc trên gel tự động được giới thiệu là BRL và Millipore.

9.5.3. Viết số liệu chuỗi mã tự động

Vì máy scanner tự động không chính xác một cách tuyệt đối, do đó người ta phải thiết kế phần mềm kèm theo, làm cho người sử dụng có thể viết ra chuỗi mã một cách tự động. Quá trình edit này rất đơn giản ngay cả trong trường hợp các đoạn phân tử chồng lấp nhau một cách ngẫu nhiên, bởi vì mỗi nucleotide được xác định rất nhiều lần. Chương trình

tổng hợp lại các chuỗi mã có thể được sử dụng để kéo thẳng ra các sequence ở những vùng chồng lấp. Máy này được sản xuất bởi Millipore và Applied Biosystems.

9.5.4. Kiểm tra số liệu

Việc kiểm tra lại số liệu của sequence phải được coi như một giai đoạn quan trọng. Computer được dùng để so sánh các sequence, tìm thấy các vùng chồng lấp, ghi nhận những khác biệt và cung cấp một biểu đồ chung về vùng chồng lấp. Thí dụ chương trình SEQED (có GCG package) báo động cho người sử dụng trường hợp có khác biệt xảy ra, giúp cho việc nhập số liệu nhanh, chính xác. Chương trình GCG-GELASSEMBLE kéo thẳng ra các sequence ở trên đỉnh của màn hình và biểu thị theo dạng giản đồ (schema) các sequence ở cuối màn hình.

9.5.5. Số lượng GC & AT

Tỉ lệ GC / AT của một sequence cho chúng ta biết đặc điểm của cấu trúc DNA hình Z và DNA hình cong. Tỉ lệ này còn cho chúng ta một dấu hiệu của một vùng mang mật mã di truyền nào đó của sinh vật. Có một vài chương trình đặc biệt như GCG's STATPLOT, cho biết số lượng GC và AT của một sequence trên biểu đồ, trong đó chúng ta tìm thấy vùng nào giàu GC hoặc AT.

9.5.6. Database của genome

Rất nhiều database của genome của một loài sinh vật nào đó đã được công bố. Nó cung cấp cho các nhà nghiên cứu phương tiện sao chụp lại hệ thống riêng của nó và sử dụng cho nhiều mục tiêu nghiên cứu khác nhau. Database của *C. elegans* ACeDB, được phát triển bởi Richard Durbin và Jean Thierry-Mieg, trở thành mẫu tiêu chuẩn về database: nhanh, dễ sử dụng và hoàn chỉnh. ACeDB sẵn sàng trong chương trình Unix, chứa tất cả các sequence của *C. elegans* từ database của EMBL DNA và bản đồ vật lý *C. elephans* của MRC Cambridge.

Một database khác AAtDB được phát triển bởi Sam Cartinhour trên *Arabidopsis thaliana*.

Một loại hình khác của database nhằm cung cấp thông tin trên internet. Đối với chương trình nghiên cứu genome của người: GDB (genetic database) được phát triển tại John Hopkins, cung cấp thông tin về bản đồ di truyền, các RFLP marker.

Sau cùng, có một vài database thuộc dạng cập nhật hóa (updated). Những database này được ghi trong text files. Thí dụ như Current *Drosophila* Genome Database (ruồi giấm) do Michael Ashburner cài đặt.

9.6. KỸ THUẬT SEQUENCING TRỰC TIẾP ĐỐI VỚI SẢN PHẨM PCR

Kỹ thuật sequencing trực tiếp đối với sản phẩm PCR là một cải tiến kỹ thuật rất quan trọng trong phân tích genome. Theo phương pháp này, chúng ta có thể biết được đặc điểm của sequence mà mình nghiên cứu một cách nhanh chóng, không cần phải xây dựng một kho lưu trữ clone (library), hoặc thanh lọc các clone.

Sản phẩm PCR có thể được chạy sequence bằng phương pháp Sanger, hay phương pháp Maxam - Gilbert.

- Sequence mục tiêu được khuếch đại, và một phần thừa ra trên dây đơn của sequence mục tiêu được tạo ra nhờ phản ứng PCR không đối xứng (asymmetric), trong đó một primer có mặt nhiều hơn gấp bội so với primer kia. Sản phẩm này được xem xét như dây template trong phương pháp Sanger.
- Người ta chuẩn bị sản phẩm PCR sao cho nó sẽ trở thành dây template thích ứng với phương pháp dideoxy trên dây đôi.
- Người ta sử dụng exonuclease để tạo ra dây đơn template từ sản phẩm PCR có dây đôi.

Về mặt nào đó, phương pháp Sanger (dideoxy) tỏ ra đơn giản hơn, nên được áp dụng phổ biến hơn. Phương pháp hóa học (Maxam-Gilbert) đòi hỏi nhiều bước thực hiện hơn, nhưng lại phù hợp với chương trình sequencing có qui mô DNA khá lớn, trong đó số lượng các sequence khác nhau cần được khuếch đại và xác định, như trong kỹ thuật có thuật ngữ là "complex sequencing" (Church và Kieffer-Higgins 1988).

9.6.1. Phương pháp dideoxy / PCR không đối xứng

Một dây đơn được khuếch đại, tạo ra nhờ thêm vào một primer có hàm lượng lớn gấp bội so với primer khác. Kết quả tạo ra sản phẩm dây đơn thặng dư này được dùng làm dây template cho phương pháp chuỗi được kết thúc bằng dideoxy, từ chuỗi đó, người ta có được sequence. Vật liệu bao gồm :

- Oligonucleotide primer 1 và 2
- ^{32}P dNTP
- 10M ammonium acetate
- ethanol loại 100 và 70
- 0,1 X TE buffer, pH 8,0
- Centricon 30 hoặc 100 cột Amicon

Điện di trên agarose gel hoặc polyacrylamide gel không có tính chất denature, sau đó, người ta cho xét nghiệm Southern blot và lai, kết tủa trong ethanol, và cho chạy sequencing, theo phương pháp dideoxy.

- Tổng hợp một PCR với các hợp phần tối ưu, và sử dụng tỉ lệ 2 nucleotide primer khoảng 100:1
- Thực hiện PCR với 40-45 chu kỳ, kết thúc bằng giai đoạn extension khá dài.
- Kiểm tra lại sản phẩm DNA dây đơn sau giai đoạn extension này, giữ cho phản ứng chậm lại trong điều kiện nhiệt độ trong phòng, rồi lấy đầu có trên mặt sản phẩm ra bằng Pasteur pipet (đầu này được nhỏ vào mẫu khi cho nó vào máy thermal cycler).
- Kết tủa bằng cách cho thêm 10M ammonium acetate đạt nồng độ sau cùng là 2,5M, theo sau đó là 1 vol rượu ethanol 100. Giữ mẫu 5 phút trong điều kiện nhiệt độ trong phòng. Cho ly tâm 5 phút, vận tốc cao, rồi rửa pellet thể rắn được kết tủa như miếng bánh trong dung môi) với ethanol 70. Sấy khô bằng cách cho ethanol bay hơi tự nhiên. Làm kết tủa pellet trở lại trong huyền phù với 50l nước hoặc 50l 0,1X TE pH 8,0.
- Thực hiện phương pháp dideoxy, sử dụng các template được tạo ra nhờ phản ứng PCR không đối xứng.

9.6.2. Phương pháp dideoxy / PCR dây đôi

Những phân tử DNA vòng, dây đôi vẫn có thể được thoái hóa (denatured) và được chạy sequence theo phương pháp dideoxy. Phương pháp chuẩn bị dây template như vậy, thường cho kết quả thấp với sản phẩm PCR dây thẳng. Do đó người ta phải cải tiến phương pháp cho thích hợp với phương pháp dideoxy, đối với DNA dây đôi được tạo ra nhờ PCR.

Phân tử DNA được khuếch đại phải thuần khiết. Primer dùng cho khuếch đại trong PCR và những primer có tính chất bên trong (internal primers) được chuẩn bị kỹ trong phản ứng sequencing.

Vật liệu thêm vào gồm có :

- Dung dịch PEG / NaCl
- Ethanol 70%, giữ lạnh
- 1 pmol (2 đến 5 ng) sequencing primer
- 5X dung dịch đệm cho annealing

Cho ly trích trong dung môi phenol / chloroform

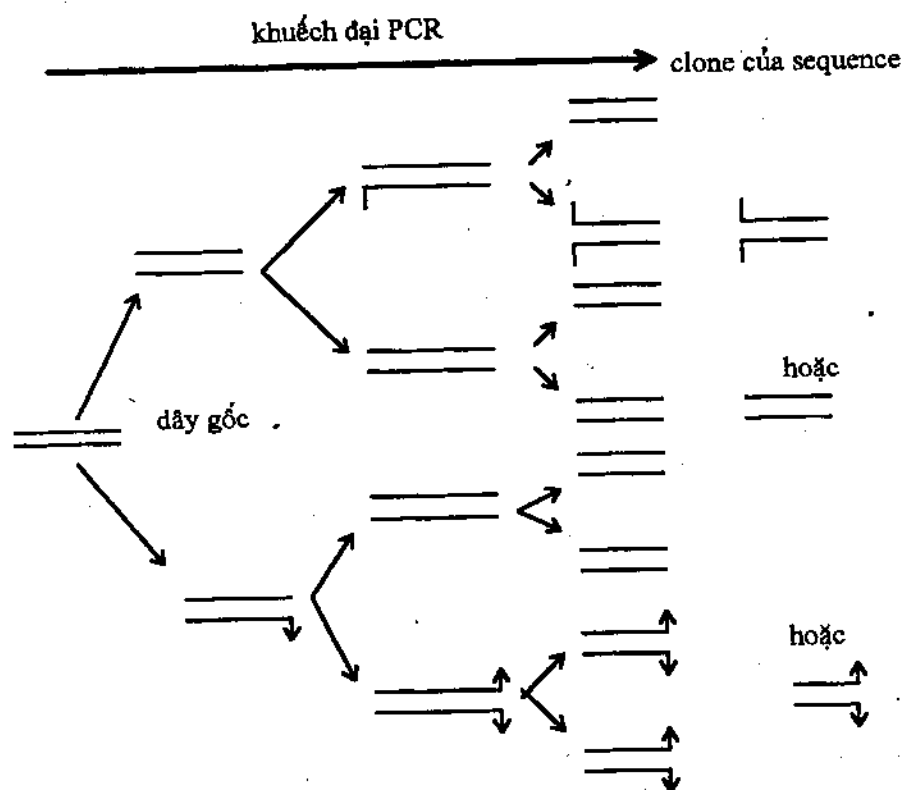
9.6.3. Phương pháp sequencing sản phẩm PCR

Church và Gilbert (1984) đã kết hợp phương pháp khuếch đại PCR và kỹ thuật sequencing lần đầu tiên. Sau đó, rất nhiều cải tiến đã và đang phát triển.

9.6.3.1. Phân tích sequence sản phẩm PCR

Tính chất trung thực của PCR là một ứng dụng quan trọng cho kỹ thuật sequencing các sản phẩm PCR. Người ta tạo ra các dòng phụ (subclone) của sản phẩm PCR, dòng này chọn lựa một phân tử trong hỗn hợp như vậy, rồi khuếch đại chúng.

Nếu mục đích của PCR là tạo ra dòng phụ DNA, thì việc xác định các vị trí giới hạn (restriction), cũng như việc khuếch đại sẽ rất khó khăn để có dây template. Do vậy, subclone là sự kiện vô cùng quan trọng trong kỹ thuật sequencing. Người ta phải chắc chắn rằng subclone có chuỗi mã riêng của nó, trong đó không có một đột biến lạ nào đó xảy ra.



Hình 9.5 : Sai sót của PCR. Hiện tượng tự tái bản DNA trong PCR có mức độ sai biệt có ý nghĩa. Sai sót nhân lên thông qua các vòng xảy ra sau đó, và sản phẩm cuối cùng là sự trộn lẫn giữa các phân tử (Alphey 1997).

Nếu mục đích của việc xác định sequence là nhằm suy luận ngược lại sequence của dây template, theo kết quả PCR trước đó, thì người ta sẽ không được suy luận sai lầm do lỗi của polymerase trong quá trình khuếch đại. Do vậy có hai cách để khắc phục vấn đề này :

- **Chạy sequence các subclone độc lập** : khuếch đại PCR một cách độc lập tránh được hiện tượng đưa vào cùng một sai sót. Do đó, phương pháp "subclone" một sản phẩm PCR từ một trong ba lần khuếch đại độc lập nhau, đã được áp dụng.

Sau đó, người ta chạy sequence các subclone này. Biến thiên của chuỗi mã giữa những subclone như vậy, có thể là nguyên nhân gây ra lỗi của polymerase. Vì những lỗi này rất hiếm xảy ra, nên bất cứ một sai lầm nào đó xuất hiện ở một trong ba subclone, người ta đều có thể suy luận ra sequence của dây template gốc do PCR tạo ra.

- **Chạy sequence sản phẩm PCR một cách trực tiếp** : phương pháp này được người ta thích áp dụng hơn. Trừ khi sai sót xảy ra ở vòng đầu tiên của quá trình khuếch đại của dây template có số lần copy rất thấp, còn lại bất cứ sai sót nào đó đều rất hiếm trong sự gộp lại các sản phẩm của phản ứng. Base tương ứng sẽ có mặt trong tư thế đa số lần át trong phân tử. Nếu sequence này xuất phát từ nhiều phân tử được xác định, thì trung bình của tất cả sequence ở vị trí nào đó sẽ là một phản ánh rất trung thực chuỗi mã của dây template đối với PCR gốc (hình 9-5).

9.6.3.2. Phát hiện thể mutant nhờ sequencing sản phẩm PCR

Cổ sở di truyền phân tử của những thể mutant đã được rất nhiều nhà di truyền nghiên cứu, đặc biệt là trong lĩnh vực y khoa, người ta cần có một xét nghiệm đáng tin cậy đối với đột biến gen gây ra bệnh có tính chất di truyền của người.

Phương pháp trước đây thường được áp dụng là xây dựng kho lưu trữ các DNA clone (library) của những gen đột biến, hoặc kho lưu trữ của các mô đột biến. Vài trường hợp, chỉ có một số lượng cực nhỏ DNA để xét nghiệm trước khi sinh. Do đó, việc khuếch đại DNA nhờ tiến bộ PCR là giải pháp được nhiều nhà nghiên cứu tán thành, với một qui mô tối hảo cho PCR và sequencing (khoảng vài trăm base), so sánh với kích thước của một gen chuẩn có ít nhất là vài kilo base.

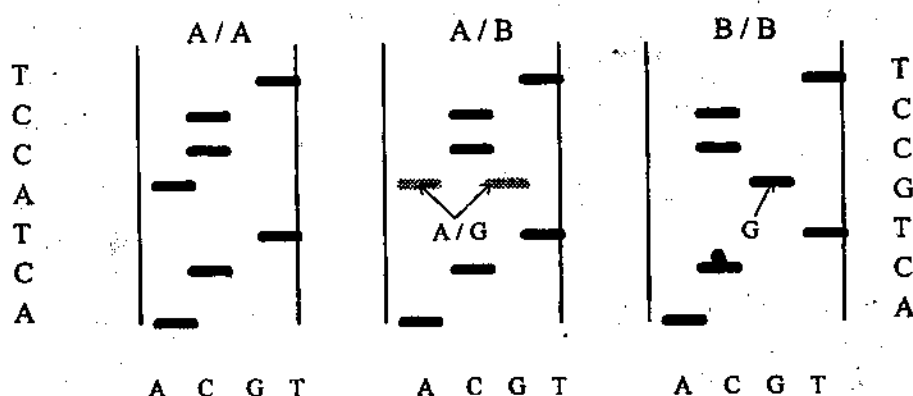
Việc phát hiện một đột biến lặn đồng hợp tử tương đối không phức tạp lắm, vì tất cả các DNA được khuếch đại đều chứa trong chuỗi mã của thể đột biến.

Kỹ thuật sequencing nhiều sản phẩm PCR gộp lại sẽ cho thấy một sequence khác ở vị trí của hiện tượng đột biến, liên quan với sequence của thể đại (wild-type). Việc xác định về định lượng đối với tỉ lệ các dạng alen càng phức tạp hơn rất nhiều.

Phát hiện và định lượng các alen khác nhau cần phải tối ưu hóa các điều kiện trong PCR để giảm thiểu đến mức thấp nhất hiện tượng khuếch đại khác biệt nhau của một thể biến dị đặc biệt nào đó (Larder và ctv. 1993). Phát hiện ra 5-methylcytosine (5-MeC) là một ví dụ của phương pháp này.

Việc phát hiện ra đột biến trội hoặc những carriers của đột biến lặn hoặc thể đa hình thì ít phức tạp hơn. Bởi vì cá thể này có hai bản sao wild-type của gen, hai bản sao thể đột biến, hoặc có mỗi một thứ trong hai. Do đó, chỉ có ba trường hợp xảy ra có thể được phân biệt ngay. Phản ứng PCR sẽ khuếch đại cả hai hiện tượng này, và sequencing sẽ cho chúng

ta biết một sự trộn lẫn giữa sequence của wild-type và sequence của thể đột biến. Trong trường hợp đột biến theo điểm (point mutation), vị trí đột biến sẽ xuất hiện ở các base của wild-type và base của thể đột biến (Hình 9.6).



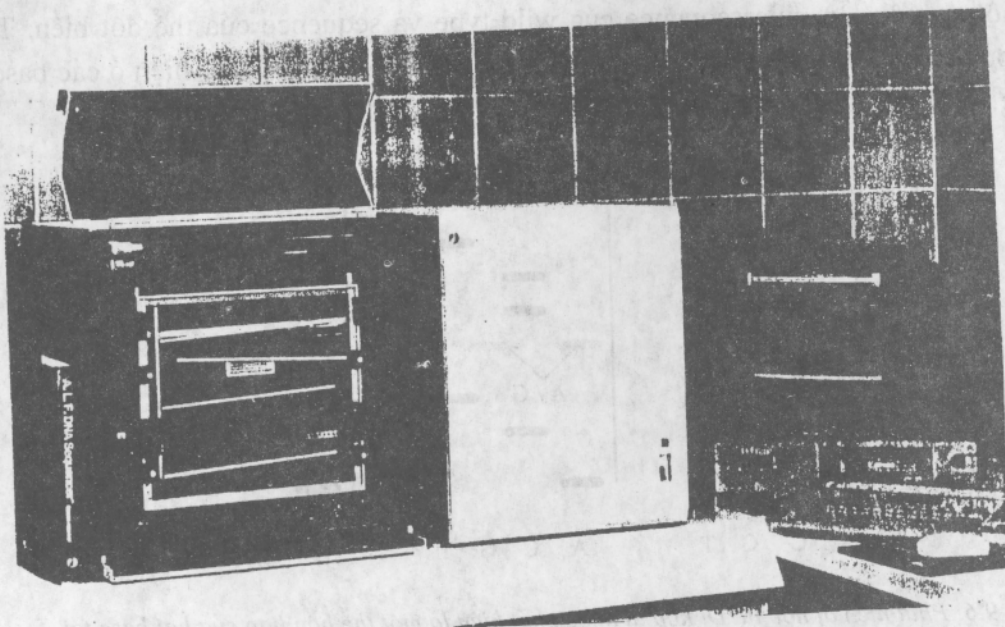
Hình 9.6: Phát hiện dị hợp tử. Dị hợp tử được phát hiện là một thể hỗn hợp của hai base tại một vị trí

9.6.3.3. Tailed primer

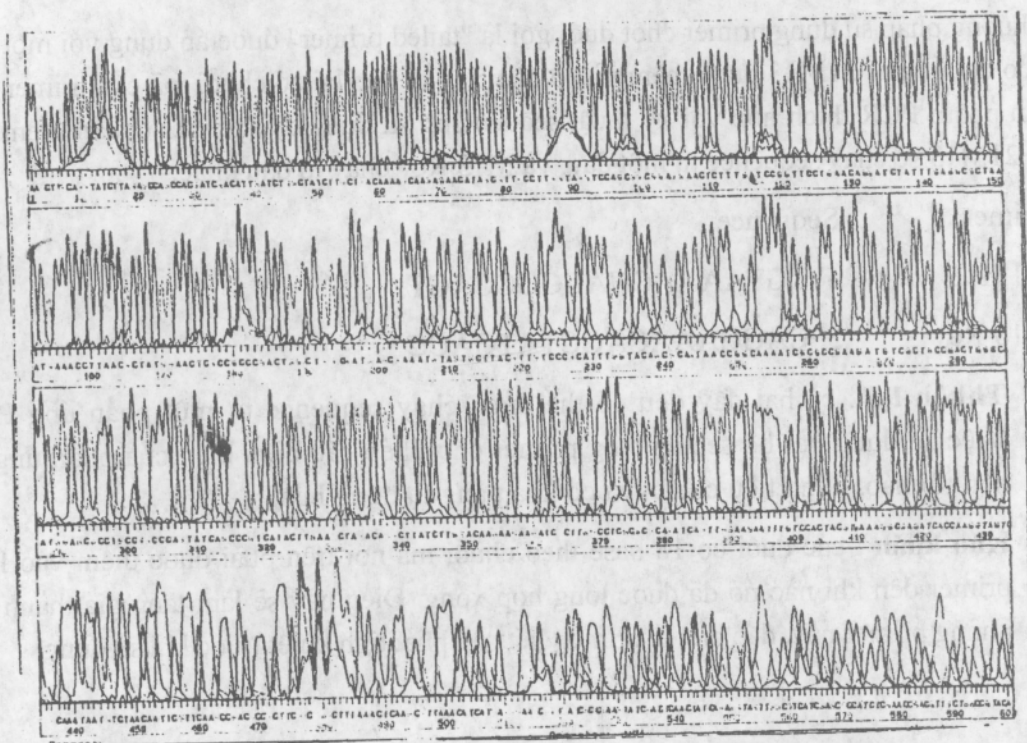
Phương pháp sử dụng primer chót đuôi gọi là "tailed primer" được áp dụng với một PCR primer có chuỗi mã -21M13 được thêm vào ở đầu 5', đầu còn lại có chuỗi mã của primer M13 Rev. Sản phẩm PCR thích hợp với kỹ thuật sequencing sử dụng dye primer chuẩn (xem mục 9.4.2) : -21M13 sẽ chạy sequence từ một đầu, và M13 Rev chạy từ đầu khác.

Primer	Sequence
-21M13	5'-TGTAACGACGGCCAGT
M13 Rev	5'-CAGGAAACAGCTATGACC

- **Thuận lợi** : cả hai dây đều có thể được chạy sequence từ một phản ứng PCR. Việc sử dụng hai bộ sequencing primer cho phép các điều kiện của phản ứng trở nên tối hảo, đúng tiêu chuẩn và có khả năng tự động hóa.
- **Khó khăn** : các đuôi có 18 base theo chuỗi mã nói trên phải được thêm vào PCR primer đến khi nào nó đã được tổng hợp xong. Điều này sẽ làm tăng giá thành của những primer này, đặc biệt là primer dài, cần phải tinh khiết trước khi sử dụng.



Hình 9.7 : Máy sequencer tự động, được vận hành bằng tia LASER huỳnh quang



Hình 9.8 : Máy đọc tự động trên computer (AutoReadIM), 100 set cho mỗi phản ứng. Mỗi đường trên đồ thị có màu khác nhau, đại diện cho một trong bốn nucleotide (A, C, G, T), chữ số nhỏ ở hàng dưới cùng biểu thị độ dài phân tử (bp)

TÀI LIỆU THAM KHẢO

1. Alphey L. 1997. *DNA Sequencing from experimental methods to bioinformatics*. BIOS Scientific Publishers, Manchester, UK. 206 p.
2. Ansorge W, B Sproat, J Stegemann, C Schwager, M Zenker. 1987. *Automated DNA sequencing: Ultrasensitive detection of fluorescent bands during electrophoresis*. Nucl. Acids Res. 15:4593-4602.
3. Chen EY, PH Seeburg. 1985. *Supercoil sequencing: A fast and simple method for sequencing plasmid DNA*. DNA(N.Y.) 4:165-170.
4. Church G, S Kiefer-Higgins. 1988. *Multiplex DNA sequencing*. Science 240:185-188.
5. Church G, W Gilbert. 1984. *Genomic sequencing*. Proc. Natl. Acad. Sci. USA 81:1991-1995.
6. Current Protocol. 1993. *DNA Sequencing*. 1993 Current Protocol. P.7.0.1- 7.7.31.
7. Hattori M, Y Sasaki. 1986. *Dideoxy DNA sequencing method using denatured plasmid templates*. Anal. Biochem. 152:232-238.
8. Larder BA, A Kohli, P Kellam, SD Kemp, M Kronick, RD Heinfrey. 1993. *Quantitative detection of HIV-1 drug resistance mutation by automated DNA sequencing*. Nature 365:671-673.
9. Maxam AM, W Gilbert. 1977. *A new method for sequencing DNA*. Proc. Natl. Acad. Sci. USA 74:560-564.
10. Middendorf LR, JC Bruce, RC Bruce. 1992. *Continuous on-line DNA sequencing using a versatile infrared laser scanner/electrophoresis apparatus*. Electrophoresis 13:487-494.
11. Sanger F, S Nicklen, AR Coulson. 1977. *DNA sequencing with chain terminating inhibitors*. Proc. Natl. Acad. Sci. USA 74: 5463-5467.
12. Smith L, J Sanders, R Kaiser, P Hughes, C Dodd, C Heiner, S Kent, L Hood. 1986. *Fluorescence detection in automated DNA sequence analysis*. Nature 321:674-679.
13. Zagursky RJ, PS Conway, MA Kashdan. 1991. *Use of ³²P for Sanger DNA sequencing*. BioTechniques 11:36-38.

Chương X

CHỌN LỌC GIỐNG CÂY TRỒNG NHỜ MARKER PHÂN TỬ DNA

Kỹ thuật di truyền và công nghệ sinh học đã tạo ra một tiềm năng to lớn cho công tác chọn tạo giống cây trồng. Việc áp dụng những kỹ thuật ở mức độ phân tử cho phép chúng ta chuyển những gen mong muốn vào trong các giống cây có cùng một loài, và chúng ta có thể du nhập những gen mới từ loài hoang dại gần gũi với loài cây trồng. Đối với những tính trạng đa gen, người ta đã gặp nhiều khó khăn trong phân tích, khi sử dụng phương pháp chọn giống cổ truyền. Nhưng điều này trở nên dễ dàng hơn, vì người ta có thể sử dụng những marker phân tử - công cụ đánh dấu rất hiệu quả. Người ta có thể hình thành nên mối quan hệ di truyền giữa những cây trồng mà trước đó nó không thể nào tương hợp nhau về giới tính. Kỹ thuật hỗ trợ cho việc chọn lọc giống thường được sử dụng là RAPD, RFLP, microsatellite, STS, AFLP (xem chương IV), hoặc ISA (inter-simple sequence repeat amplification) và ALP (amplicon length polymorphism) trên quần thể F₂ hoặc quần thể hồi giao, quần thể các dòng đẳng gen (NIL), quần thể đơn bội kép (DH), và quần thể các dòng cận giao tái tổ hợp (RIL).

Vấn đề có tính chiến lược trong ứng dụng di truyền phân tử vào lĩnh vực chọn giống chính là : "**chọn giống nhờ marker phân tử**", với thuật ngữ thông dụng MAS, viết tắt từ chữ "marker-assited selection". Những thể đánh dấu phân tử (molecular tags) là điều kiện đầu tiên trong chiến lược MAS, đã được phát triển trong nhiều loài cây trồng, với nhiều loại hình khác nhau của marker phân tử. Những marker phân tử có nhiều ưu điểm hơn các marker hình thái cổ điển. Do đó nó rất có ích cho các nhà chọn tạo giống. Hiệu quả cải tiến giống cây trồng sẽ gia tăng gấp nhiều lần so với chọn giống cổ điển, nhờ thực hiện việc chọn lọc không cần trực tiếp trên tính trạng mong muốn, mà thông qua marker phân tử liên kết với tính trạng đó. Điều này đòi hỏi marker phải liên kết rất chặt với tính trạng mong muốn. Hơn nữa, marker này phải không bị điều tiết bởi môi trường, không bị ảnh hưởng bởi điều kiện trong đó cây trồng đang sinh trưởng và phát triển.

Trong phân tích genome, thường có bốn mục tiêu được đặt ra (Paterson 1996) :

- (1) Xác định nơi ký sinh gây hại cây trồng và bảo vệ giống cây trồng.
- (2) Đánh giá mức độ đa dạng về di truyền.
- (3) Du nhập gen mới từng bước (introgression) hoặc hồi giao.
- (4) Lập bản đồ các tính trạng.

Như vậy trong chiến lược MAS, mục tiêu 3 và 4 sẽ được thực hiện, và chúng ta sẽ thảo luận các nội dung này trong chương X.

10.1. HỒI GIAO

Mục tiêu chung của công tác chọn giống cây trồng là làm sao chuyển đổi một gen (một tính trạng) đặc biệt nào đó vào trong một kiểu gen (giống) mong muốn. Phương pháp hồi giao đã được đề nghị để đáp ứng yêu cầu này. Thông thường trong ngân hàng gen cây trồng, nó chỉ thể hiện được một phần về mức độ đa dạng di truyền trong loài của nó. Nhiều gen kháng sâu bệnh hại, kháng điều kiện đất đai, thời tiết bất lợi, có mặt ở các loài hoang dại, có quan hệ gần với giống cây trồng. Nhưng các loài hoang dại thường thể hiện những liên kết gen điều khiển tính trạng nông học xấu, bên cạnh gen mục tiêu. Khả năng của phương pháp hồi giao đã được thẩm định trong lịch sử chọn giống. Nó được xem như phương tiện rất có hiệu quả để khống chế những liên kết bất lợi như vậy. Hiện nay, với phương tiện của marker phân tử DNA, người ta càng phát triển tiềm năng của hồi giao trong nghiên cứu và ứng dụng việc du nhập gen từ các loài hoang dại, thẩm định khả năng introgression từ loài hoang dại sang loài cây trồng (Paterson 1996). Đối với con lai trong phương pháp hồi giao, Tanksley và Rick (1980) đã có thể chọn lọc nhờ isozyme marker. Con lai được ghi nhận tính chất introgression với số thế hệ hồi giao ít hơn so với phương pháp cũ. Bản đồ di truyền của marker phân tử DNA còn phong phú hơn isozyme marker rất nhiều lần, do đó hiệu quả chắc chắn cao hơn. Nghiên cứu trên cây lúa, người ta đã du nhập thành công gen kháng rầy nâu từ lúa hoang *Oryza australiensis* vào lúa trồng. Hơn 600 cây BC₂F₄ đã được phân tích isozyme (Multani và ctv. 1994). Sau đó, Ishii và ctv. (1994) đã dùng RFLP để phân tích giữa dòng introgression và dòng tái tục (recurrent), con lai F₂, F₃. Kết quả họ đã xác định gen kháng rầy nâu *Bph-10* đối với biotype 3 đã được du nhập từ lúa hoang sang lúa trồng, gen này định vị trên nhiễm thể 12, liên kết với marker RG457, với khoảng cách di truyền là $3,68 \pm 1,29$ cM.

Phương pháp hồi giao được trợ giúp nhờ DNA marker đã và đang được ứng dụng rộng rãi, tiết kiệm 1-2 thế hệ, giúp cho việc cải tiến cây trồng đáp ứng với sản xuất sớm hơn rất nhiều (Paterson 1996). "**Hồi giao nhờ marker phân tử**" được thực hiện trong việc chuyển nạp vào cây trồng những gen lạ, bằng phương tiện công nghệ sinh học.

Theo nguyên tắc, những DNA marker có thể thúc đẩy sự hồi phục của giống tái tục. Bằng cách sử dụng DNA marker để xác định các thể "tái tổ hợp" [recombinant-], người ta thấy rằng: các đoạn của nhiễm thể lạ được du nhập vào, có thể được sắp xếp một cách trật tự trong một không gian chật hẹp như vậy. Giống tái tục bị đột phá bởi những alen không mong muốn, liên kết rất chặt với tính trạng mục tiêu (Tanksley và Crick 1980).

Nhiều nghiên cứu gần đây cho thấy: hiện tượng du nhập ngẫu nhiên (random introgression) các chromatin lạ vào giống cây trồng, người ta ghi nhận sự biến dị có tính

chất khó hiểu (cryptic) gia tăng - biến dị không phải có tính chất tạm thời của kiểu hình từ nguồn hoang dại. Thí dụ như giống cà chua hoang dại *Lycopersicon pennellii*, phát triển chậm, có chứa một đoạn nhiễm thể đặc biệt điều khiển tính trạng gia tăng mức độ tích lũy trọng lượng khô ở giống cà chua bình thường, một tính trạng mà nó thường do kiểu hình của cây quyết định (Paterson 1996). Phương pháp **hồi giao ngẫu nhiên** của chromatin lại vào trong giống triển vọng, trong nghiên cứu về biến dị khó hiểu, là phương pháp **hồi giao cải tiến** (advanced-backcross). Thuật ngữ này được viết tắt "(AB)-QTL analysis". Nó được ứng dụng khá thành công đối với cà chua và bông vải, và hiện được nhiều nhà chọn giống cây trồng khác quan tâm.

10.2. BẢN ĐỒ DI TRUYỀN

Kỹ thuật mapping (lập bản đồ) và sequencing (viết chuỗi mã) của genome cây trồng sẽ giúp cho chúng ta rất nhiều về việc làm sáng tỏ chức năng của gen, sự điều tiết của gen và sự thể hiện của gen. Bản đồ liên kết gen có chất lượng cao hiện được thực hiện ở nhiều loài cây trồng. Những marker phân tử đã được dùng để xác định và đánh dấu các gen mong muốn (Mohanty và ctv. 1997). Phân tích liên kết gen là một trong những phương pháp căn bản, cần thiết trong nghiên cứu di truyền. Liên kết gen có thể khẳng định khoảng cách di truyền giữa những tính trạng đa hình, được ghi nhận khác nhau về sự có mặt của các enzyme hoạt động, sự thể hiện chiều dài hạn chế của đoạn phân tử, hoặc chuỗi mã nucleotide ở tại một locus nào đó. Bản đồ liên kết gen trên cơ sở marker hình thái và isozyme marker đã được công bố trên cây lúa, bắp, lúa mì, lúa mạch và nhiều cây trồng khác. Trong cây lúa, 185 marker đã được xác định trên 12 nhiễm thể (Kinoshita 1995). Cho dù những thông tin của bản đồ di truyền cổ điển này rất quan trọng, giúp ta biết được vị trí của gen tương ứng với các tính trạng về kiểu hình. Nhưng tính chất hữu dụng của nó rất hạn chế, bởi vì số marker hình thái còn quá ít, không thỏa mãn yêu cầu của nhiều chương trình chọn giống. Bên cạnh đó, sự thể hiện của những marker hình thái bị ảnh hưởng bởi điều kiện môi trường. Điều này làm cho marker hình thái kém sự thu hút so với marker phân tử DNA trong cải tiến giống cây trồng.

Tính đa hình của chuỗi mã nucleotide thường đủ để chúng ta kết luận về chức năng của marker phân tử trong kỹ thuật mapping. Tính chất đa hình này (polymorphism) được minh họa rõ nét trong kỹ thuật phân tử, thí dụ như RFLP, AFLP, SSR, RAPD, STS, v.v.. Bản đồ di truyền được hình thành trên nhiều loài cây trồng, sử dụng các marker nói trên, trong một quần thể phân ly nhất định nào đó.

Trong số các marker phân tử DNA, marker RFLP được sử dụng đầu tiên trong việc lập bản đồ gen của con người (Botstein và ctv. 1980), sau đó nó được cải biên để ứng dụng cho kỹ thuật mapping cây trồng. RFLP là thể đa hình đáng tin cậy nhất, có thể được dùng cho những phân tích chính xác về kiểu gen. RFLP có tính chất codominant, có thể xác định

một locus đồng nhất, xác định cả alen trội và lặn. Khi những cDNA có chức năng về gen khá rõ ràng, được dùng làm marker, thì vị trí của gen đặc biệt nào đó cần tìm sẽ được xác định cụ thể trên nhiễm thể. Sử dụng RFLP marker, bản đồ di truyền đã được người ta phát triển trên nhiều loài cây trồng, đặc biệt là trên cây lúa. Tính đa hình được phát hiện bởi RFLP marker sẽ rất đáng tin cậy vì nó bao gồm sự xác nhận của enzyme phân cắt hạn chế, và sự lai với thể thăm dò (pre) chuyên tính nào đó. Bản đồ di truyền của lúa sử dụng 2300 DNA marker, trong đó có 1500 cDNA marker do chương trình genome cây lúa của Nhật Bản (RGP) tạo ra (Kurata và ctv. 1994). Causse và ctv.(1994) cũng đã phát triển bản đồ di truyền cây lúa, bằng cách sử dụng 800 RFLP.

Bảng 10-1: Các marker phân tử

Marker	Thuật ngữ bằng tiếng Anh	Tác giả
AFLP	Amplified fragment length polymorphism	Vos et al. 1995
ALP	Amplicon length polymorphism	Ghareyazie et al. 95
AP-PCR	Arbitrarily primed PCR	Welsh et al. 1990
AS-PCR	Allele specific PCR	Sarkar et al. 1990
CAPS	Cleaved amplified polymorphis sequences	Lyamichev et al. 93
DAF	DNA amplification fingerprinting	Caetano et al. 1991
ISA	Inter-SSR amplification	Zietkiewicz et al. 94
RAPD	Random amplified polymorphic DNA	William et al. 1991
RFLP	Restriction fragment length polymorphism	Botstein et al. 1980
SAP	Specific amplicon polymorphism	William et al. 1991
SCAR	Sequence characterized amplified region	William et al. 1991
SSCP	Single strand conformation polymorphism	Orita et al. 1989
SSLP	Microsatellite simple sequence length polymorphism	Salentijn et al. 1995
SSLP	Minisatellite simple sequence length polymorphism	Jarman et al. 1989
SSR	Simple sequence repeats	Hearne et al. 1992
STS	Sequence tagged sites	Fukuoka et al. 1994

Tuy nhiên RFLP yêu cầu một công việc có số lao động rất lớn, và nhiều thời gian. Các phương pháp cải tiến gần đây đều dựa trên tiến bộ kỹ thuật của PCR (xem chương VIII). Thí dụ như RAPD được phát triển bởi Welsh và McClelland (1990), William và ctv (1990). Cho đến nay, có nhiều marker trên cơ sở PCR được cải tiến, đã được phát triển cho kết quả rất triển vọng (bảng 10-1).

William và ctv. (1991) đã chuyển đổi RFLP marker thành những chuỗi mã có tính chất vùng khuếch đại (SCAR), nhờ kỹ thuật sequencing hai đầu các dòng DNA của genome, và thiết kế các oligo-primer trên cơ sở các chuỗi mã hai đầu như vậy. Người ta sử dụng các primer này một cách trực tiếp trên DNA của genome, thông qua phản ứng PCR, nhằm khuếch đại vùng đa hình. Nếu tính chất đa hình của độ dài đoạn phân tử không được khuếch đại, chúng ta phải nhờ các enzyme phân cắt hạn chế trên sản phẩm PCR, để phát hiện RFLP nào có trong đoạn phân tử khuếch đại này. SCAR có ưu điểm là: codominant giúp ta ghi nhận cả gen trội và lặn, trong khi RAPD chỉ có tính chất dominant, không ghi nhận tốt trong trường hợp gen lặn.

AFLP được áp dụng trên sở khuếch đại PCR của những đoạn phân tử hạn chế, phát sinh bởi enzyme phân cắt chuyên tính nào đó, và nó dựa trên cơ sở các adapter (là oligo nucleotide). Phương pháp này tạo ra một lượng lớn các băng trong điện di, làm cho việc phát hiện thể đa hình thuận lợi hơn rất nhiều. Số đoạn phân tử DNA khuếch đại có thể được kiểm soát, bằng cách chọn lọc base khác nhau và thành phần nucleotide trong adapter. Mặc dù chưa có bản đồ di truyền nào được phát triển hoàn chỉnh bằng AFLP marker, cho đến nay, nhưng phương pháp này ngày càng được áp dụng phổ biến để phát triển các marker đa hình. Bản đồ liên kết gen, sử dụng AFLP marker, gần đây đã được thực hiện trên cây lúa mạch (barley) (Becker và ctv. 1995). Kỹ thuật này rất có ích để hoàn chỉnh bản đồ và phân nhóm di truyền giống cây trồng. Lin và ctv. (1996) đã so sánh RFLP, RAPD và AFLP trong khi nghiên cứu tính đa hình của đậu nành, họ nhận xét rằng: AFLP là kỹ thuật có hiệu quả nhất để tìm kiếm thể đa hình. Nhờ tính chất tái sản xuất cao, tần suất cao về tính đa hình đồng nhất, AFLP trở thành phương pháp phân tích DNA có hiệu quả để xác định đa hình và xác định liên kết gen, bằng cách phân tích cá thể trong quần thể phân ly. Tuy nhiên AFLP vẫn còn đắt tiền, vì phải nhuộm bằng điện di bằng bạc, fluorescent, hoặc hoạt động chất phóng xạ.

SSR hay microsatellite marker phản ánh thể đa hình trên cơ sở số đơn vị lặp lại của một vùng mục tiêu nào đó của genome (Jacob và ctv. 1991, Litt và Luty 1989, Weber và May 1989). Người ta đã thiết lập bản đồ di truyền trên chuột với hơn 6000 SSR, khoảng cách trung bình là 1,1 cM (Dietrich và ctv. 1996). Số lượng và thành phần của microsatellite thay đổi trên thực vật so với động vật. Tần suất lặp lại chuỗi mã dài hơn 20 bp, xảy ra trên động vật là 33 kb, trong khi trên động vật có vú là 6 kb (Wang và ctv. 1994). Ở di truyền của người, AC và TC là những đơn vị lặp lại rất phổ biến, nhưng ở thực vật thì AT phổ biến hơn, theo sau đó là AG hoặc TC (Powell và ctv. 1996). Nói chung là ở thực vật, nó có SSR ít hơn 10 lần so với người. Chuỗi mã nucleotide kế cận chuỗi mã lặp lại được dùng để thiết kế primers nhằm khuếch đại số lượng khác nhau của đơn vị lặp lại trong những giống khác nhau. Loại đa hình này có tính chất tái sản xuất rất cao (highly reproducible). Những primer như vậy thực sự rất cần cho yêu cầu phát hiện nhanh và chính

xác các loci đa hình. Thông tin đó có thể được dùng để phát triển bản đồ vật lý trên cơ sở những sequence tags như vậy. Hiện chúng ta chưa có nhiều các primer để phát hiện SSRP trên thực vật so với genome của người.

SSCP marker là công cụ rất mạnh và nhanh, nhưng nó chỉ áp dụng cho việc tìm kiếm thể đa hình của những đoạn phân tử DNA tương đối ngắn (Orita và ctv. 1989). SSCP có thể xác định tính chất dị hợp tử của những đoạn phân tử DNA (có cùng trọng lượng phân tử), và nó có thể phân biệt được sự thay đổi của một vài nucleotide nào đó. Người ta cho nó là công cụ hữu dụng trong xét nghiệm bệnh di truyền ở người. Trong thực vật, SSCP chưa được phát triển nhiều. Người ta hi vọng nó sẽ giúp cho việc phân nhóm di truyền con lai trở nên dễ dàng hơn, khi chúng ta có primer thích hợp đối với tính trạng quan trọng nào đó (Fukuoka và ctv. 1994).

10.3. PHÂN TÍCH QTL

Hầu hết các tính trạng nông học quan trọng như năng suất, yếu tố hình thành năng suất, ngày trở hoa, v.v.. được kiểm soát bởi đa gen. Hoạt động đa gen và sự tương tác lẫn nhau, hiện vẫn còn được biết rất ít trong khi chúng điều khiển tính trạng số lượng. Trên cây cà chua, người ta đã áp dụng RFLP marker để phân tích sự cải tiến các tính trạng số lượng (Osborn và ctv. 1987, Tanksley và Hewitt 1988, Paterson và ctv. 1991). Các yếu tố cải tiến tính trạng số lượng (QTMF = quantitative trait modifying factors) cũng được xác định như những loci ảnh hưởng đến tính trạng số lượng thông qua epistasis. Trong khi đó, QTL có ảnh hưởng kiểu hình tương đối lớn và độc lập, trên cơ sở di truyền (Mohan và ctv. 1997). Marker phân tử có liên kết chặt với gen điều khiển độ chắc của quả cà chua, gen phục hồi sự thụ tinh, và gen tương hợp rộng (wide compatibility), gen kháng hạn (Nguyễn và ctv. 1997), hình thái rễ, tính trạng rụng hạt của cây lúa, và hiện tượng apomixis của *Pennisetum squamulatum* đã được công bố (Mohan và ctv. 1997). Đối với cây lúa, các tính trạng chiều dài hạt cơm khi nấu chín, tính kháng đạo ôn không hoàn toàn, tính chống chịu bệnh đốm vằn, ngày trổ, chiều dài thân đã được áp dụng QTL marker để phân tích. Bản đồ liên kết gen với mật độ dày cũng đã được thiết lập với sự trợ giúp của phần mềm MAPMARKER / QTL (Yano và ctv. 1995).

Thông thường, người ta rất khó xác định vị trí chính xác và hoạt động gen của các loci tính trạng số lượng. Để khắc phục nhược điểm trên, người ta thành lập các nguồn vật liệu di truyền (genetic stocks) có tính trạng được biết rõ ràng, thí dụ như các dòng gần đẳng gen (NIL), mang một hoặc nhiều đoạn nhiễm thể của một dòng bố mẹ, trên nền tảng di truyền của một dòng bố mẹ khác. Người ta thực hiện công tác điều tra trên toàn genome để xác định đoạn nhiễm thể được du nhập vào, bằng phương tiện RFLP. Nhờ sử dụng các dòng NIL đối với một đoạn nhiễm thể nào đó, người ta có thể xử lý một QTL nào đó như một yếu tố đơn trong di truyền Mendel. Bản đồ RFLP có chất lượng cao, có mật độ dày các

marker, sẽ giúp cho công tác xác định dễ dàng hơn. Nó còn giúp cho chúng ta lập được bản đồ và xử lý các tính trạng số lượng thành các yếu tố di truyền Mendel. Phương pháp này bao gồm : (i) tìm kiếm sự kết hợp giữa những RFLP marker đang phân ly, và (ii) xác định sự ích lợi của một quần thể phân ly để tìm thấy liên kết của marker với QTL.

Trên cây cà chua, các dòng có tính chất bổ sung (substitution lines) và chồng lấp nhau đã được dùng để lập bản đồ QTL (Nienhuis và ctv. 1987). Nhờ vậy các vị trí chính xác của một QTL nào đó trên bản đồ liên kết gen đều được xác định. Khi bản đồ các gen có lợi được thành lập với giải pháp cao, các gen bấy giờ có thể trở thành mục tiêu cho kỹ thuật cloning dựa trên cơ sở bản đồ, mà thuật ngữ chuyên môn thường dùng là "**map-based cloning**". Những vị trí có tính chất giả định và marker phân tử DNA liên kết với QTL cũng sẽ làm cho MAS trở nên dễ dàng hơn trong chọn giống cây trồng.

10.4. BẢN ĐỒ VẬT LÝ

Người ta yêu cầu thiết lập bản đồ vật lý nhằm đáp ứng (i) phân lập gen của tính trạng mong muốn, và (ii) định tính nó. Do đó, các công cụ như YAC, BAC, cosmid đã được khai thác rất có hiệu quả để cloning các phân tử DNA có kích thước lớn từ 40 kb đến 600 kb, hoặc lớn hơn. Mỗi vectơ đều có giá trị riêng của nó. Mặc dù người ta cần có số lượng clone ít hơn để tạo ra contig đối với một vùng mục tiêu nào đó, khi sử dụng YAC, nhưng người ta phải chịu một hiện tượng "khảm" (chimerism) trong đoạn phân tử DNA được gắn vào vectơ. Trái lại, cosmid tỏ ra thuận lợi hơn trong chuyển nạp để sản xuất các cây transgenic, nhưng nó lại bị hạn chế về khả năng mang DNA với kích thước nhỏ, làm cho chúng ta rất khó thiết lập hiện tượng chồng lấp trên một vùng rộng lớn của genome. BAC có ưu điểm hơn YAC là có ít thể khảm hơn, cho dù nó mang DNA có kích thước không bằng YAC. Bên cạnh đó, BAC còn có ưu điểm là dễ xử lý hơn, mức độ sinh sản của vi khuẩn nhanh hơn men (yeast).

Kho lưu trữ (library) YAC cho lúa, *Arabidopsis*, cà chua; và kho lưu trữ BAC cho *Arabidopsis*, cao lương, và lúa đã được thiết lập (Mohan và ctv. 1997). Người ta đã và đang khai thác các kho lưu trữ này, thiết lập các contig, thiết lập bản đồ vật lý. Có thể nói hiện nay bản đồ có chất lượng cao nhất đối với thực vật là bản đồ của *Arabidopsis*. Hơn 90% nhiễm thể số 4, và 80% nhiễm thể số 2 đã được bao phủ bởi YAC. Khoảng 65%, 60% và 85% nhiễm thể số 1, 3, và 5 cũng được bao phủ bởi YAC contig, theo thứ tự (Mohan và ctv. 1997). Trên cây lúa, có khoảng 1200 RFLP và STS marker được sử dụng để thanh lọc kho lưu trữ YAC, bao gồm 7000 clone, có kích thước gắn vào trung bình là 350 kb. Theo kết quả này, khoảng 2200 YAC độc lập đã được xác định và xếp thứ tự. Chương trình RGP của Nhật đã bao phủ hơn 50% genome cây lúa bằng YAC này. Đặc biệt là YAC contig trên nhiễm thể số 6 đã được công bố rất đầy đủ ((Umehara và ctv. 1995).

10.5. ÁP DỤNG BẢN ĐỒ DI TRUYỀN TRONG CHỌN TẠO GIỐNG

Người ta xác định có 3 vấn đề làm cho sự tiến triển của công tác cải tiến giống cây trồng trở nên chậm lại trong những năm gần đây: (1) thiếu biến dị di truyền mới (gen / QTL) do chúng ta ít sử dụng nguồn gen từ bên ngoài (exotic germplasm); (2) không có khả năng chọn lọc kiểu hình giữa các alen mong muốn và không mong muốn, hoặc kiểu gen có tính chất nhiều loci đối với các kiểu hình phức tạp bao gồm nguồn gen từ bên ngoài; (3) khả năng hạn chế của nhà chọn giống trong việc tìm nguồn gen và ghi nhận các thông tin di truyền cần thiết bằng nỗ lực riêng của họ (Li và ctv. 1997). Theo phân tích đa dạng di truyền bằng isozyme, người ta nhận thấy có khoảng 50% mức độ đa dạng của cây lúa *Oryza sativa* phân bố trong quần thể, 25% có ở giữa các loài phụ, và 25% có ở giữa các quần thể của *O. sativa* (Li và Rutger 1998). Người ta tự hỏi: trong vật liệu di truyền đầu tiên được sử dụng để cải tiến cây trồng, mức độ đa dạng đã được khai thác như thế nào? Các thành tựu về kỹ thuật DNA marker trong những năm gần đây, đã cung cấp cho chúng ta nhiều công cụ rất hiệu quả, đánh dấu nhiều bước phát triển quan trọng trong lĩnh vực di truyền, chọn giống, tiến hóa và nghiên cứu genome (Tanksley và McCouch 1997). Thành công nhất trong kỹ thuật này là phát triển bản đồ liên kết gen phân tử trong nhiều loài sinh vật. Từ đó nó làm nền tảng cho các nghiên cứu khác như bản đồ QTL, chọn tạo giống nhờ marker, bản đồ vật lý, và kỹ thuật map-based cloning các gen quan trọng về kinh tế (Li 1998).

Một ứng dụng quan trọng khác của bản đồ liên kết gen là tạo điều kiện để nhà nghiên cứu phân tích ở qui mô phân tử (molecular dissection) các tính trạng phức tạp thông qua thiết kế, xem xét các thí nghiệm QTL mapping. Nó kết hợp với nhiều kiểu lai phân tích khác nhau (lai ba thử nghiệm để tìm epistasis, trắc nghiệm Scaling, trắc nghiệm liên kết digenic, v.v..) và những phương pháp phân tích thống kê khác nhau. Kỹ thuật DNA marker cho phép chúng ta mô tả tính chất di truyền của những tính trạng phức tạp này, trong trường hợp mọi chi tiết không thể dự đoán trước (Paterson 1997). Nghiên cứu lập bản đồ QTL sẽ cung cấp cho chúng ta 3 đặc điểm sau đây về di truyền số lượng (Li 1997) :

- Trong những cặp lai, các tính trạng hình thái học hoặc sinh lý như chiều cao, ngày trổ bông được xác định bởi một vài QTL marker, mỗi marker sẽ có ảnh hưởng kiểu hình tương đối lớn. QTL có ảnh hưởng lớn như vậy thường được tìm thấy trong kỹ thuật mapping trên những quần thể và môi trường khác nhau.
- Các tính trạng số lượng như năng suất và yếu tố hợp thành năng suất thường được kiểm soát bởi các loci có tương tác không alen (epistasis) và QTL có ảnh hưởng tương đối nhỏ.
- Những QTL có ảnh hưởng nhỏ và loci tương tác không alen này, thường được ghi nhận: có tương tác kiểu gen x môi trường (G x E) và ảnh hưởng của chúng thường

chịu sự thay đổi về nền tảng di truyền, không thể dự đoán được. Do đó, tương tác $G \times E$ và epistasis sẽ kết hợp với QTL làm cho nó trở nên khó khăn hơn trong ứng dụng QTL-MAS để cải tiến di truyền của những tính trạng phức hợp này. Thí nghiệm để tìm hiểu cận kề tương tác QTL x môi trường rất hiếm bởi vì nó đắt tiền và tốn nhiều thời gian.

10.6. QUAN HỆ GIỮA CHỌN GIỐNG CỔ ĐIỂN VÀ CHỌN GIỐNG NHỜ MARKER

Sự kết hợp trong phân tích thống kê giữa những alen tại những loci của marker phân tử và những alen tại các loci có tính chất số lượng (QTL), có thể được sử dụng để chọn lọc giống một cách gián tiếp, nhưng có khả năng chính xác rất cao (Stubber và ctv. 1982, Soller và Beckmann 1983). Khi người ta chọn lọc các tính trạng có hệ số di truyền thấp, như năng suất chẳng hạn, nhà chọn giống thường sử dụng hình thức chọn lọc theo gia đình để nâng cao hệ số di truyền một cách có hiệu quả. Rõ ràng là người ta sẽ có ít cơ hội hơn trong việc gia tăng hiệu quả chọn lọc, bằng phương pháp sử dụng thông tin di truyền phân tử. Chúng ta nên xem xét ở một mức độ rộng hơn về chiến lược lai tạo giống cây trồng, bao gồm qui mô (độ lớn) gia đình có thể được, phù hợp với phương pháp chọn giống nhờ marker (MAS) trong các gia đình nhỏ, hoặc ở mức độ cá thể từng cây. Điều này sẽ cho ra kết quả tốt hơn, ít tốn công sức hơn, so với phương pháp cổ điển.

Phương pháp phân tích thống kê đối với MAS, trên cơ sở marker di truyền phân tử thông qua phương pháp đo lường kiểu hình chuẩn, đều nhằm mục đích: có được chỉ số chọn lọc tối hảo.

10.6.1. Liên kết gen và số loci của marker

Nguồn của hiện tượng không cân bằng trong liên kết gen (linkage disequilibrium) là (i) sự chuyển dịch gen một cách ngẫu nhiên, (ii) sự lai giữa hai dòng khác nhau về di truyền.

Sự chuyển dịch gen một cách ngẫu nhiên (random genetic drift) thường làm gia tăng hiện tượng không cân bằng trong liên kết gen, giữa một cặp loci. Gọi mức độ tái tổ hợp của chúng là r . Người ta ghi nhận r sẽ nhỏ hơn $1 / (4N_e)$. N_e là độ lớn quần thể Wright có hiệu quả (Hill và Rerton 1968).

Theo sau đó là sự kiện lai giữa hai dòng cận giao, trong một quần thể lớn "t" thế hệ sau thế hệ F_2 hoặc S_1 . Sự xuất hiện không cân bằng trong liên kết gen được dự đoán, giữa một cặp loci có r ít hơn 10cM, sẽ là e^{-rt} trong điều kiện giao phối ngẫu nhiên, hoặc $e^{-2r(1-1/2)t}$, trong điều kiện tự thụ hoàn toàn.

Ngoại trừ trong một quần thể có độ lớn hữu hiệu rất nhỏ, sự lai là nguồn gốc có cơ chế mạnh hơn sự chuyển dịch gen ngẫu nhiên, để tạo ra hiện tượng không cân bằng trong liên kết gen. Tuy nhiên, sự tái tổ hợp sẽ làm mất dần ảnh hưởng không cân bằng trong liên kết gen ban đầu do sự kiện lai tạo ra. Trong một genome, độ dài tổng cộng của bản đồ liên

kết gen là L Morgan, số marker phân tử tối thiểu để phát hiện ra hiện tượng không cân bằng trong liên kết gen với QTL là $2L + c$ trong ngẫu giao (random mating), hoặc $4L (1 - 1/2^c) + c$ trong tự phối hoàn toàn. c là yếu tố đại diện cho số đơn bội (giao tử) của nhiễm thể. Xem thí dụ ở bảng 10-2.

Bảng 10.2 : Số loci marker phân tử tối thiểu cần thiết để phát hiện sự không cân bằng trong liên kết gen với QTL ở các thời điểm khác nhau sau khi lai hai dòng cận giao của cây nhị bội, tứ bội, và lục bội

Độ dài bản đồ (cM)	Số nhiễm thể	Hệ thống lai tạo	Thế hệ kể từ khi lai		
			1	5	10
1000	10	lai chéo	30	110	210
2000	20	lai chéo	60	220	420
3000	30	lai chéo	90	330	630
1000	10	tự thụ	30	49	50
2000	20	tự thụ	60	98	100
3000	30	tự thụ	90	146	150

Sự chọn lọc tự nhiên hay nhân tạo đối với các phối hợp gen có tính chất epistasis cũng có thể sản xuất hiện tượng không cân bằng trong liên kết gen. Sự chọn lọc nhân tạo trên một phối hợp giữa thông tin phân tử và kiểu hình được người ta hi vọng rằng sẽ làm giảm sự kết hợp giữa những loci của marker và QTL bởi hiện tượng cố định các alen của marker, và hiện tượng thay đổi linkage disequilibria. Hoạt động tái tổ hợp và chọn lọc rất cần cho việc đánh giá lại sự phối hợp giữa những alen tại các marker loci và QTL trong mỗi thế hệ (Lande 1996).

10.6.2. Chọn lọc cá thể trong quần thể tạp giao

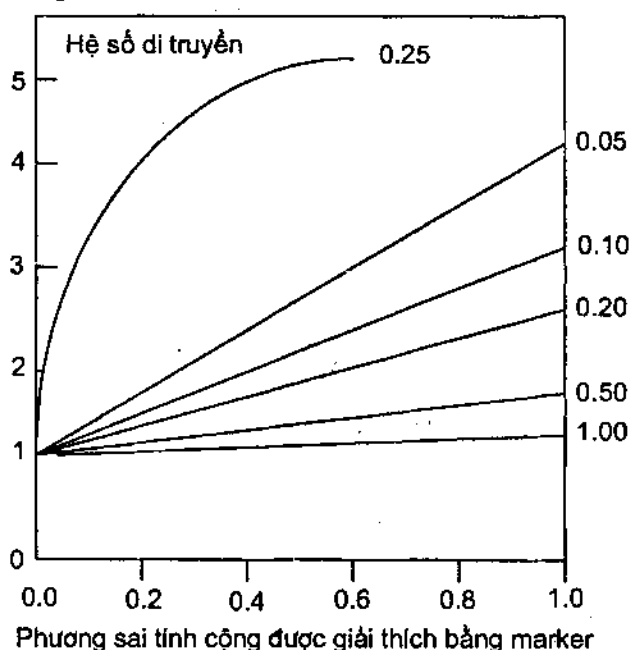
Nhằm làm sáng tỏ những nguyên tắc cơ bản của MAS, chúng ta hãy xem xét việc chọn lọc cá thể để cải tiến một tính trạng nào đó trong quần thể cây thụ phấn tự do, hoặc thế hệ F₂, hoặc S₁ của một cặp lai giữa hai dòng cận giao. Giả dụ rằng, các kiểu hình của cá thể và trạng thái có tính chất alen tại một số loci của marker phân tử được ghi nhận bằng cách lấy mẫu trong quần thể.

Từ những giả định cổ điển về sự thay đổi nhỏ của tần suất gen ở mỗi thế hệ, đáp ứng chọn lọc theo kiểu hình được căn cứ trên các ảnh hưởng cộng tính (additive) của QTL. Sự phát triển của nó được gắn với hiện tượng biến dị tại những loci của marker phân tử. Sự phát triển này có thể được ước đoán bằng phương trình đa tuyến của giá trị kiểu hình cá thể

"z" trên số bản sao (0, 1, hoặc 2) của những alen đặc biệt tại loci của marker có tính đa hình. Đối với những loci của marker có tính chất diallelic, người ta có thể chọn một alen một cách ngẫu nhiên để cho vào hệ phương trình đa tuyến này. Với nhiều alen, người ta có thể cho vào tất cả theo kiểu phân biệt riêng ra (Falconer 1981, Kempthorne 1969). Khi bản đồ liên kết của marker loci sẵn sàng, người ta có thể hoàn tất công việc lập hệ thống phương trình đa tuyến riêng biệt nhau, đối với từng nhiễm thể. Bởi vì, hiện tượng không cân bằng trong liên kết gen giữa các loci thường nhỏ, trên các nhiễm thể khác nhau.

Phương pháp dùng phương trình đa tuyến giải thích phần nào hiện tượng không cân bằng trong liên kết giữa những marker loci, giữa những marker, và giữa những QTL mà nó liên kết với nhau. Điều này cho phép thông tin di truyền từ nhiều hơn một marker locus trên QTL được sử dụng. Nó rất quan trọng trong việc tối đa hóa phương sai di truyền của tính cộng (additive) đối với tính trạng nào đó, được giải thích bằng marker phân tử. Bởi vì, sự tái tổ hợp của bất cứ marker đơn nào dường như chỉ có một tương quan không hoàn toàn với QTL mà nó liên kết (Lande 1996). Hoàn thiện hệ phương trình đa tuyến riêng biệt đối với mỗi nhiễm thể sẽ giúp chúng ta tránh được các sai phạm về mặt thống kê, trong số những alen có tính chất độc lập của marker.

Hiệu quả tương đối của MAS



Hình 10.1 : Hiệu quả chọn lọc nhờ marker trên những cá thể, sử dụng chỉ số chọn lọc tối hảo kết hợp với số liệu kiểu hình và số liệu phân tử, so với chọn lọc cổ điển.

Hệ số gốc trong phương trình đa tuyến sẽ ước lượng "ảnh hưởng tính cộng" tạm thời, đối với tính trạng được kết hợp bởi những alen tại các marker loci. Những hệ số gốc này có thể được qui tụ vào một thang điểm có tính chất hệ thống phân tử (net molecular score), gọi

là "m", đối với mỗi cá thể. Nó là giá trị tổng số của những ảnh hưởng cộng (additive) phối hợp với tất cả alen của marker (Lande và Thompson 1990).

Kiểu hình "z" và thang điểm hệ thống "m" của một cá thể có thể cho vào một chỉ số chọn lọc tuyến tính (linear selection index)

$$I = b_z z + b_m m \quad (1)$$

trong đó b_z và b_m (giá trị tương đối) được chọn để tối đa hóa mức độ cải tiến kiểu hình trung bình. Bởi vì m không có giá trị kinh tế thực sự, cho nên (1) có thể viết:

$$b_m / b_z = (I / h^2 - 1) / (1 - p) \quad (2)$$

trong đó h^2 là hệ số di truyền của một tính trạng, và p là tỉ lệ của phương sai di truyền do tính cộng, được giải thích bởi marker loci. Như vậy thang điểm hệ thống trong chỉ số chọn lọc được tính một cách chặt chẽ đối với các tính trạng có hệ số di truyền thấp, trong khi hiệu quả của tính cộng khá cao (Lande và Thompson 1990).

Hiệu quả của MAS đối với chỉ số chọn lọc này có thể được giải thích theo kiểu chọn lọc kiểu hình của phương pháp cổ điển, có cùng một cường độ chọn lọc. Giả sử z, là những giá trị có phân bố chuẩn (normal distribution), hiệu quả tương đối trong cùng một cường độ chọn lọc có thể được tính: (Smith 1967, Lande và Thompson 1990)

$$[p / h^2 + (1 - p)^2 / (1 - h^2 p)]^{1/2} \quad (3)$$

Phần đầu tiên giải thích $(p / h^2)^{1/2}$ là hiệu quả chọn lọc, chỉ có trên những marker, liên quan đến một kiểu hình nào đó. Phần còn lại thể hiện hiệu quả cộng tính được tạo ra từ kiểu hình trong chỉ số chọn lọc dọc theo thang điểm hệ thống phân tử. Đối với các tính trạng có hệ số di truyền thấp, nó có thể được xem xét như sau: p ở mức so sánh thấp nhất so với h^2 , sự tăng thêm hiệu quả ngay sau đó có thể đạt được nhờ vào MAS.

Hiệu quả tối đa, trong trường hợp $p = 1$, sẽ là $1 / h$ (Hình 10-1). Đối với các tính trạng có hệ số di truyền thấp, hiệu quả của MAS sẽ rất lớn, so với phương pháp chọn lọc giống cổ điển. Riêng tính trạng có hiệu quả cộng tính càng cao, càng dễ kết hợp với các loci của marker phân tử.

10.6.3. Phương sai di truyền về tính cộng được giải thích bằng marker

Khả năng phát hiện QTL tùy thuộc vào điều kiện có thể được của marker loci, trong sự không cân bằng của liên kết, mà chúng ta đã thảo luận ở phần trên. Mức độ đóng góp của phương sai di truyền của tính cộng đối với một tính trạng nào đó, được giải thích nhờ marker loci, cũng lệ thuộc vào ảnh hưởng của QTL và độ lớn của mẫu quần thể (số cá thể) được sử dụng để phân tích.

Mỗi QTL chỉ tham gia vào một phần nhỏ của biến dị kiểu hình nói chung, và số marker loci có liên kết gần với tính trạng nghiên cứu, có đủ để xem xét phương sai cộng tính tại một QTL đặc biệt nào đó. QTL này được phát hiện ở mức độ α về ý nghĩa thống kê, nếu mức độ đóng góp phương sai di truyền của tính cộng lớn hơn $4x^2 / h^2N$ (Lande và Thompson 1990). Ở đây, N là độ lớn của mẫu, và x_α là số độ lệch chuẩn trên giá trị trung bình của phân bố chuẩn (normal), sao cho mức độ α có ý nghĩa ($x_{0,05} = 1,65$ và $x_{0,01} = 2,33$). Người ta cần những mẫu lớn để tìm kiếm QTL của những tính trạng có hệ số di truyền thấp.

Biến dị di truyền của những tính trạng số lượng thường được tạo ra bởi một vài loci có ảnh hưởng tương đối lớn và nhiều loci khác có ảnh hưởng nhỏ hơn (Shrimpton và Rertson 1988, Paterson và ctv. 1988). Một mô phỏng đã được phát triển theo phân bố hình học của ảnh hưởng QTL, để phân tích mức độ đóng góp của phương sai do tính cộng (Lande và Thompson 1990). Từ đó, người ta có thể phát hiện ra hàng loạt các marker loci có ích trong một mẫu mà độ lớn được biết trước. Giả sử nhiều QTL có ảnh hưởng lớn nhất không liên kết, hoặc gần như cân bằng liên kết với QTL khác. Nếu mỗi QTL đều đóng góp có mức độ vào phương sai tính cộng với giá trị theo tỉ số " a ", thì số yếu tố di truyền số lượng có thể được định nghĩa là: $n_E = (1 + a) / (1 - a)$. Đây là mô hình giống như của Wright (1968), về số yếu tố hữu hiệu, nhưng dựa trên cơ sở công thức phương sai di truyền tính cộng, nhiều hơn là di truyền quần thể (Lande 1996).

Giá trị thực sự của n_E biến thiên từ 2 đến 10, và hệ số di truyền thấp $h^2 = 0,1$. Trong trường hợp này, độ lớn của mẫu: " N " cần phải có từ vài trăm đến vài nghìn cá thể cho marker loci giải thích phương sai tính cộng. Số mẫu phải lớn khi n_E lớn và hệ số di truyền thấp. Phân tích này dựa trên giả định lấy mẫu ngẫu nhiên của những cá thể mà số liệu về kiểu hình và số liệu về marker đã được biết.

Chúng ta cố gắng tránh sử dụng chỉ số chọn lọc trong MAS, bởi vì mức độ sai lệch (bias) sẽ khá lớn, do có quá nhiều thông số đè nặng trên sự diễn dịch thông tin ở mức độ phân tử và làm suy giảm hiệu quả chọn lọc.

10.6.4. Chọn lọc gia phả

Phương pháp chọn giống theo phả hệ là phương pháp phổ biến trong cải tiến giống cây trồng. Việc ứng dụng MAS trong những cây cá thể sẽ giúp cho chúng ta chọn lọc một cách chính xác, và cung cấp cho chúng ta một công cụ rất mạnh để giám định ở thế hệ F_1 , F_2 mà việc chọn lọc bằng gặp nhiều hạn chế.

MAS có thể là một phương tiện rất hữu hiệu trong trường hợp chọn lọc những tính trạng quá khó đo đếm, hoặc rất đắt tiền khi đo đếm, như tính kháng hạn, kháng lạnh, hoặc một số loại côn trùng gây hại. Chúng ta có thể sử dụng các marker loci và QTL có liên kết với những tính trạng kháng hoặc chống chịu ấy, phân tích trên F_2 . Việc tuyển chọn bằng

marker như vậy có thể được thực hiện trong một môi trường bình thường (không có hạn và sâu hại), vào giai đoạn cây còn nhỏ. Chi phí cho sự đánh giá như vậy tùy thuộc vào loại marker mà chúng ta sử dụng. Người ta có xu hướng tìm kiếm các marker ngày càng rẻ và dễ thực hành với qui trình đơn giản hơn.

Ngay cả đối với những thế hệ sau của chọn lọc gia phả, việc chọn lọc tập trung vào những giá trị di truyền có tính chất phân lập rõ ràng, MAS sẽ cho chúng ta cơ hội tốt để gia tăng hiệu quả chọn lọc.

Đối với cây tự thụ phấn, thí dụ chúng ta có khả năng trắc nghiệm 50.000 cây, và chọn được 50 cây tốt nhất. Sau đó chúng ta có thể xem xét tình trạng số cây trên một dòng (1-1000) và số dòng cận giao sẽ biến thiên 50 - 50.000. Với số cây cụ thể trên một dòng, nếu chúng ta tăng số cây / dòng có nghĩa là gia tăng hệ số di truyền trong những dòng, nhưng làm giảm đi cường độ chọn lọc (Lande 1996). Do đó người ta nghiên cứu số cây tối hảo trên một dòng nhằm tối đa hóa sự đáp ứng của việc chọn lọc. Đó là vấn đề có tính chất kinh điển trong lý thuyết chọn lọc. Số cây tối hảo trên một dòng tùy thuộc vào tổng số cây được trắc nghiệm, số dòng được chọn, và hệ số di truyền của tính trạng trong những cây cá thể của các dòng cận giao khác nhau. Thông tin về marker phân tử của một chỉ số chọn lọc đối với MAS, có thể làm thay đổi số cây tối hảo trên một dòng nào đó.

Chúng ta thử xem xét một quần thể của những dòng cận giao. Giả thiết rằng các biến thiên trong các dòng là kết quả có tính chất môi trường và có tính chất phát triển, bên cạnh đó không có khác biệt có ý nghĩa về môi trường giữa các dòng. Gọi h^2 là hệ số di truyền của tính trạng trong những cây cá thể (một cây / dòng). Hệ số di truyền của kiểu hình trung bình của dòng là H^2 với n cây trên một dòng:

$$H^2 = nh^2 / [(n-1)h^2 + 1] \quad (A1)$$

Tỉ số phương sai giữa các dòng (với n cây trên dòng) đối với phương sai cá thể trong các dòng là v^2 :

$$v^2 = h^2 + (1 - h^2) / n \quad (A2)$$

Chỉ số chọn lọc tối hảo đối với MAS giữa các dòng cận giao có cùng một dạng là chỉ số đối với việc chọn lọc cá thể trong phương trình (1) và (2), nhưng H^2 lại chứa đựng trong h^2 . Gọi p là tỉ số của phương sai di truyền tổng cộng giữa các dòng được giải thích bằng marker loci. Phản ứng dự đoán đối với MAS trong các dòng có n độ lớn là:

$$i\sqrt{H^2} [p / H^2 + (1 - p)^2 / (1 - H^2 p)]^{1/2} \quad (A3)$$

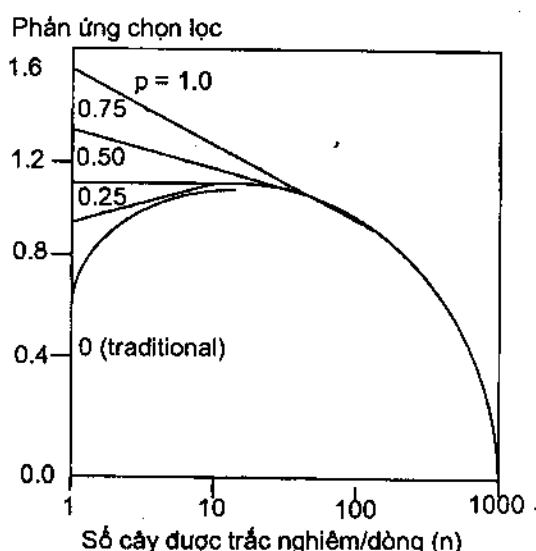
trong đó i là hằng số chọn lọc chuẩn khác nhau giữa các họ. Đối với một tính trạng có phân bố chuẩn giữa các họ, phương sai đơn vị $f(x)$ [unit variance], thì i tương trưng cho chiều cao đường cong biểu diễn tại một điểm cắt nào đó $f(y)$, chia cho giá trị tương xứng được lưu giữ trong chọn lọc Q . Trong một quần thể có N dòng cận giao, và n cây trong mỗi

dòng, nếu chúng ta biết được tổng số cây được đo, và số dòng được chọn, thì $Q = k / N$. Gọi k dòng được chọn từ nN cây. Nếu $nN = k \times 1000$ cây, thì $Q = k/n = n/1000$ và

$$i = 1000 f(y) / n \quad (A4)$$

Đối với bất cứ một giá trị nào của n tương ứng với giá trị của Q , điểm cắt tại y và chiều cao $f(y)$ có thể được tìm thấy trong bảng phân bố chuẩn (normal distribution). Trong hình 10-2, chúng ta thảo luận trong trường hợp $k=15$, nhưng kết quả hoàn toàn độc lập đối với k ($k>1$).

Hình 10-2 còn cho thấy trong điều kiện chọn lọc kiểu hình truyền thống, sự đáp ứng chọn lọc gần như tối đa, ghi nhận từ 5 đến 20 cây trên một dòng. Tại điểm có độ lớn gia đình là 20 hoặc lớn hơn, sẽ có ít cơ hội để cải tiến nhờ MAS.

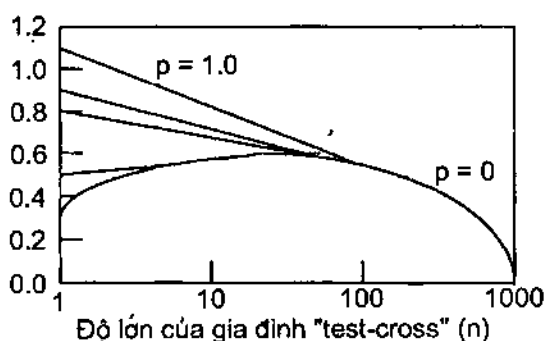


Hình 10.2 : Phản ứng trên thế hệ chọn lọc giữa các dòng cận giao. Giả sử $h^2 = 0,2$ và tổng số cây được trắc nghiệm là 50.000, với 50 dòng tốt nhất được chọn (Lande 1996)

10.6.5. Tuyển chọn tái tục

Dạng nguyên thủy của tuyển chọn tái tục (recurrent selection) đối với năng suất giống ưu thế lai như cây bắp (*Zea mays* L.) là một ví dụ, xảy ra trong những chu kỳ của ba thế hệ (Hull 1945). Ở thế hệ đầu tiên, những cá thể trong quần thể được lai với dòng "tester", rồi cho tự thụ. Ở thế hệ thứ hai, cây bố mẹ được tuyển chọn trên cơ sở ưu việt của con lai. Ở thế hệ thứ ba, người ta đem trồng những hạt tự thụ từ bố mẹ được tuyển chọn, rồi cho lai chéo để tạo ra bố mẹ cho chu kỳ tiếp theo. Chọn lọc tái tục có tính chất thuận nghịch (reciprocal recurrent selection) bao gồm công tác cải tiến đồng thời các dòng ưu thế lai ưu việt giữa hai quần thể, mẫu được chọn của quần thể này sẽ là tester của quần thể kia (stock và ctv. 1949). Phương pháp tuyển chọn tái tục là phương pháp có hiệu quả lâu dài trong chọn giống (Hallauer 1981).

Chúng ta thử phân tích một chương trình tuyển chọn tái tục với một tester có tính chất cận giao (inbred tester) trong một môi trường đơn tính. Đây là chương trình nghiên cứu dài hạn về năng suất của cây bắp tại Đại Học Iowa State (Hallauer và ctv. 1988). Ngày nay nó chính là phương pháp chọn lọc tái tục thuận nghịch, nhưng kết quả của phân tích được cải tiến nhờ bổ sung phương pháp MAS. Giả sử có 15.000 cây lai được trắc nghiệm trong mỗi chu kỳ, và 15 bố mẹ tốt nhất được chọn. Hệ số di truyền về năng suất cây cả thể là $h^2 = 0,2$ và độ lớn của gia đình được lai thử nghiệm (test cross) là 50 trong môi trường đơn tính, sao cho có 300 cây bố mẹ được trắc nghiệm trong mỗi chu kỳ. Kết quả trong mỗi chu kỳ là một hàm số theo độ lớn của gia đình test-cross, sử dụng chỉ số chọn lọc đối với MAS (xem hình 10-3).



Hình 10.3 : Phản ứng chọn lọc tái tục / chu kỳ với một tester cận giao theo n, đối với các giá trị p khác nhau (1, 0.75, 0.50, 0.25 và 0). Biết rằng $h^2 = 0,2$, số con lai là 15.000, số bố mẹ tốt nhất được chọn lọc là 15. Phản ứng chọn lọc được tính theo độ lệch chuẩn kiểu hình của quần thể bố mẹ lai chéo.

Tính trạng được chọn lọc trong quần thể bố mẹ lai chéo là kết quả của con lai trong những cặp lai có một tester cận giao (đồng hợp tử). Kiểu hình của con lai có thể do một phần tương tác tính trội (dominance interaction) giữa những alen của bố mẹ lai chéo và tester cận giao này. Nếu không xem xét epistasis hoặc coi như nó không có, thì phương sai di truyền trong quần thể bố mẹ lai chéo là kết quả hoạt động của gen cộng tính, bởi vì bất cứ ở một locus nào đó của con lai chỉ chứa đựng một alen từ một bố mẹ, ngược lại với cơ sở di truyền của dòng tester cận giao. Giả sử QTL trong quần thể bố mẹ theo qui luật cân bằng của Hardy-Weinberg và sự cân bằng trong liên kết, với một số lượng rất to lớn con lai trong mỗi lần lai thử nghiệm, thì một nửa giá trị phương sai di truyền tổng cộng sẽ thuộc về "giữa các bố mẹ", và một nửa thuộc về "trong bố mẹ".

Hệ số di truyền của kết quả chung với n con lai trong mỗi lần lai thử nghiệm được ký hiệu là H^2 :

$$H^2 = nh^2 / [(n - 1)h^2 + 2] \quad (A5)$$

Đối với các tính trạng mà phương sai di truyền của bố mẹ rất có ý nghĩa về tính cộng, thí dụ như năng suất cá thể của cây bắp (Hallauer 1981), h^2 có thể xem như hệ số di truyền của tính trạng của cá thể trong quần thể bố mẹ.

Chỉ số chọn lọc tối hảo đối với tuyển chọn tái tục nhờ marker phân tử có cùng một dạng trong phương trình thử nghiệm (1) và (2) nhưng với H^2 thay vì h^2 và p là tỉ số của phương sai di truyền xảy ra trong con lai được giải thích bằng marker. Phản ứng dự đoán đối với MAS giữa những bố mẹ dựa theo cơ sở kết quả trung bình của n con lai trên mỗi lần lai thử nghiệm, nó được tính theo đơn vị độ lệch chuẩn (căn bậc hai của phương sai kiểu gen + phương sai do môi trường).

$$i [hH / (2)^{1/2}] [p / H^2 + (1 - p)^2 / (1 - H^2 p)]^{1/2} \quad (A6)$$

Nếu tổng số cây lai là $k \times 1000$ được trắc nghiệm và số k tốt nhất được chọn lọc, thì i có cùng công thức (A4).

10.6.6. Thảo luận mối quan hệ giữa chọn giống cổ điển và chọn giống nhờ marker

Phương pháp chọn giống cổ điển như chọn gia phả, chọn dòng tái tục bao gồm việc chọn kiểu hình giữa các dòng cận giao, hoặc giữa các dòng bố mẹ lai chéo trên cơ sở lai thử nghiệm (test-cross). Đối với những tính trạng có hệ số di truyền thấp như năng suất chẳng hạn, người ta sẽ có một khoảng biến thiên tối hảo của độ lớn gia đình, làm tối đa hóa phản ứng trên một thế hệ, hoặc trên một chu kỳ chọn lọc. Với độ lớn gia đình trong chương trình cải tiến giống cây trồng, biến thiên từ 15 đến 30 cá thể trên môi trường, hệ số di truyền thường có giá trị trung bình hoặc cao, sẽ tạo ra một cơ hội rất thấp để gia tăng hiệu quả chọn lọc, thông qua việc áp dụng marker di truyền phân tử.

Tuy nhiên, nếu hiện tượng mất cân bằng trong liên kết đủ giữa những "marker loci" và QTL, nhờ lai giữa các dòng cận giao với nhau, hoặc do hiện tượng chuyển dịch gen ngẫu nhiên (random genetic drift), thì những marker phân tử có thể đóng góp vào rất lớn phương sai di truyền của tính trạng, làm gia tăng một cách có hiệu quả hệ số di truyền trong những cá thể. Với điều kiện như vậy, MAS đối với các cá thể (hoặc đối với gia đình), thông qua sử dụng chỉ số chọn lọc kết hợp với đánh giá phân tử và đánh giá kiểu hình, có thể sẽ làm gia tăng **hiệu quả chọn lọc** cao hơn so với phương pháp chọn giống cổ truyền (xem hình 10-2 và 10-3, đường cong $p = 0$ là phương pháp cổ truyền).

MAS còn tạo nên các cơ hội để gia tăng **cường độ chọn lọc** trong những thế hệ đầu tiên của phương pháp gia phả, trong khi phương pháp cổ truyền gần như không có chọn lọc hoặc chọn lọc bằng mắt rất yếu để loại bỏ các dòng xấu nhất. MAS có thể được xem là đặc biệt hữu dụng trong phương pháp gia phả đối với những tính trạng rất khó và rất đắt tiền để thanh lọc, quan sát con lai, như tính kháng hạn, tính kháng sâu bệnh hại, tính kháng lạnh, tính quang cảm nhờ nhiệt độ trong lúa lai, v.v.. (Lande 1996, Lang 1998). Kết hợp giữa marker loci và QTL sẽ giúp chúng ta xem xét tính kháng hạn hoặc sâu hại ngay ở thế hệ F_2 .

Ở những thế hệ sau đó, việc tuyển chọn có thể được hoàn tất trên cây con (mạ) hoặc cây trưởng thành, trồng ở môi trường bình thường, giảm bớt số dòng trắc nghiệm ở môi trường có vấn đề.

Lande (1996) đã cung cấp một số liệu ước lượng rằng: đối với cây trồng hằng năm tại Mỹ, chi phí cho một cây (trung bình của nhiều loài khác nhau) trong chương trình chọn giống cổ điển là 0,05 đến 0,50 \$US. Trong khi đó, chi phí cho chương trình chọn giống nhờ marker phân tử (MAS) biến thiên từ 0,10 đến 1,00 \$US / marker locus / cây. Điều này tùy thuộc vào loại marker được sử dụng.

Sự kết hợp giữa chọn giống cổ điển và MAS là chiến lược đang được nhiều quốc gia xem xét và vận dụng một cách hợp lý, để nâng cao hiệu quả chọn lọc giống phục vụ cho yêu cầu sản xuất.

TÀI LIỆU THAM KHẢO

1. Becker J, P Vos, M Kuiper, F Salamini, M Heun. 1995. *Combined mapping of AFLP and RFLP markers in barley*. Mol. Gen. Genet. 249:65-73.
2. Botstein B, RL White, M Skolnick, RW Davis. 1980. *Construction of a genetic linkage map in man using restriction fragment length polymorphism*. Am J Hum Genet 32:314-331.
3. Causse MA, TM Fulton, YG Cho, SN Ahn, J Chunwongse, K Wu, J Xiao, Z Yu, PC Ronald, SE Harrington, G Second, SR McCouch. 1994. *Saturated molecular map of the rice genome based on an interspecific backcross population*. Genetics 138:1251-1274.
4. Comstock RE, HF Rinson, PH Harvey. 1949. *A breeding program designed to make maximum use of both general and specific combining ability*. Agronomy Journal 41:360-367.
5. Dietrich WF, Miller J, R Steen, MA Merchant, D Damron-Boles, Z Husain, R Dredge, MJ Daly, KA Ingalls, TJ O'Connor, CA Evans, MM DeAngelis, DM Levinson, L Kruglyak, N Goodman, NG Copeland, NA Jenkins, TL Hawkins, L Stein, DC Page, ES Lander. 1996. *Comprehensive genetic map of the mouse genome*. Nature 380:149-152.
6. Falconer DS. 1981. *Introduction to quantitative genetics*. 2nd ed. Longman, London.
7. Hallauer AR, WA Russell, KR Lamkey. 1988. *Corn breeding*. In: Corn and corn improvement. Agronomy Monograph no. 18, 3rd ed. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, pp. 463-564.
8. Hallauer AR. 1981. Selection and breeding methods. In: Frey KE (ed), *Plant Breeding II*. Iowa State University Press, Ames, pp. 3-35.
9. Hill WG, R Robertson. 1968. *Linkage disequilibrium in finite populations*. Theor. Appl. Genet. 38:226-231.
10. Hull FH. 1945. *Recurrent selection for specific combining ability in corn*. Journal of the American Society of Agronomy 37:134-145.
11. Ishii T, DS Brar, DS Multani, GS Khush. 1994. *Molecular tagging of genes for brown planthopper resistance and earliness introgressed from Oryza australiensis into cultivated rice O. sativa*. Canadian Journal of Genome 37(2):217-221.

12. Jacob HJ, K Lindpaintner, SE Lincoln, K Kusumi, RK Bunker, YP Mao, D Ganten, VJ Dzau, ES Lander. 1991. *Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat*. Cell 67:213-224.
13. Kempthorne O. 1969. *An introduction to genetic statistics*. Iowa State University Press, Ames.
14. Kinoshita T. 1995. *Report of committee on gene symbolization, nomenclature and linkage group*. Rice Genet. Newsl. 12:9-153.
15. Kurata N, G Moore, Y Nagamura, T Foote, M Yano, Y Minobe, M Gale. 1994. *Conservation of genome structure between rice and wheat*. Biotechnology 12:276-278.
16. Lande R, R Thompson. 1990. *Efficiency of marker-assisted selection in the improvement of quantitative traits*. Genetics 124:743-756.
17. Lande R. 1996. *Marker-assisted selection in relation to traditional methods of plant breeding*. P.437-451. Chapter 20. University of Oregon Press, Eugene.
18. Lang NT. 1998. *Development of PCR-based markers for thermosensitive genetic male sterility tms-3(t) in rice*. 102 p. Post-doctoral Dissertation. IRRI, Los Banos, Philippines.
19. Li ZK, SRM Pinson, AH Paterson, WD Park, JW Stansel. 1997. *Epistasis for three grain yield components in rice (Oryza sativa L.)*. Genetics 145:453-465.
20. Li ZK. 1997. *Molecular analysis of epistasis affecting complex traits*. In Molecular Analysis of Complex Traits. Ed. by AH Paterson, CRC Press, Boca Raton, New York.
21. Li ZK. 1998. *An integrated international rice molecular breeding program*. Outline 15p. IRRI. Philippines.
22. Lin JJ, J Kuo, J Ma, JA. Saunders, HS Beard, MH MacDonald, W Kenworthy, GN Ude, BL Matthews. 1996. *Identification of molecular markers in soybean: comparing RFLP and AFLP DNA mapping techniques*. Plant Mol. Biol. Rep. 14:156-169.
23. Litt M, JA Luty. 1989. *A hypervariable microsatellite revealed by in vitro amplification of a nucleotide repeat within the cardiac muscle actin gene*. Am. J. Hum. Genet. 44:397-401.
24. Mohan M, S Nair, A Bhagwat, TG Krishna, M Yano. 1997. *Genome mapping, molecular markers and marker-assisted selection in crop plants*. Molecular Breeding 3:87-103.
25. Multani DS, K Jena, DS Brar, BG de los Reyes, ER Angeles, GS Khush. 1994. *Development of monosomic alien addition lines and introgression of genes from Oryza*

- australiensis* Domin. to cultivated rice *O. sativa* L. Theor. Appl. Genet. 88:102-109.
26. Nguyen HT, CR Babu, A Blum. 1997. *Breeding for drought resistance in rice: Physiology and Molecular Genetics Considerations*. Crop Sci. 37:1426-1434.
 27. Nienhuis J, T Helentris, M Slocum, B Ruggero, A Shaefer. 1987. *Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato*. Crop Sci. 27: 797-803.
 28. Orita M, H Iwahana, H Kanazawa, K Hayashi, T Sekiya. 1989. *Detection of polymorphism of human DNA by gel electrophoresis as single strand conformation polymorphisms*. Proc. Natl. Acad. Sci. USA 86:2766-2770.
 29. Osborn TC, DC Alexander, JF Fobes. 1987. *Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit*. Theor. Appl. Genet. 73:350-356.
 30. Paterson AH, ES Lander, JD Hewitt, S Paterson, SE Lincoln, SD Tanksley. 1988. *Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphism*. Nature (London) 335:721-726.
 31. Paterson AH, SD Tanksley, ME Sorrells. 1991. *DNA markers in plant improvement*. Adv. Agronomy 46:39-90.
 32. Paterson AH. 1996. *Genome mapping in plants*. Texas A&M Univ. Acad. Press. RG Landes Co. Austin, USA. 330 p.
 33. Powell W, GC Machray, J Provan. 1996. *Polymorphism revealed by simple sequence repeats*. Trends Plant Sci. 1:215-222.
 34. Shrimpton AE, A Robertson. 1988. *The isolation of polygenic factors controlling bristle score in Drosophila melanogaster*. II. Distribution of the third chromosome bristle effects within chromosome sections. Genetics 118:445-459.
 35. Smith C. 1967. *Improvement of metric traits through specific genetic loci*. Animal Production 9:349-358.
 36. Soller M, JS Beckmann. 1983. *Genetic polymorphism in varietal identification and genetic improvement*. Theor. Appl. Genet. 67:25-23.
 37. Stuber CW, MM Goodman, RH Moll. 1982. *Improve of yield and ear number resulting from selection at allozyme loci in a maize population*. Crop Sci. 22:737-740.
 38. Tanksley SD, CM Rick. 1980. *Isozyme gene linkage map of the tomato: application in genetics ang plant breeding*. Theor. Appl. Genet. 57:161-170.
 39. Tanksley SD, J Hewitt. 1988. *Use of molecular markers in breeding for soluble solids*

content in tomato - a re-examination. Theor. Appl. Genet. 75:811-823.

40. Umehara Y, A Inagaki, H Tanoue, Y Yasukochi, Y Nagamura, S Saji, Y Otsuki, T Fujimura, N Kurata, Y Mine. 1995. *Construction and characterization of a rice YAC library for physical mapping. Molecular Breed.* 1:79-89.
41. Weber JL, PE May. 1989. *Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. Am J Hum Genet* 44:388-396.
42. Welsh J, M McClelland. 1990. *Fingerprinting genomes using PCR with arbitrary primers. Nucl. Acids Res.* 8:7213-7218.
43. William JGK, AR Kubelik, KJ Livak, JA. Rafalski, SV Tingey. 1990. *DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res.* 18:6531-6535.
44. Wright S. 1968. *Evolution and the genetics of populations. Vol. I. Genetic and biometric foundations.* University of Chicago Press, Chicago.
45. Yano M, Y Nagamura, N Kurata, T Sasaki. 1995. *Genetic linkage map of DNA markers and its application to genetic analysis in rice. Gamma Field Symposia (Institute of Radiation Breeding, NIAR/MAFF, pan)* p.34.

THUẬT NGỮ CHUYÊN MÔN

Abundance: của một mRNA là số trung bình các phân tử có trong tế bào

Acentric: một nhiễm thể không bình thường, hoặc một đoạn nhiễm thể thiếu tâm động (centromere) trong quá trình phân bào

Acrocentric: một nhiễm thể có tâm động của nó rất gần với điểm cuối của nhiễm thể, tạo cho nó có một vai rất ngắn so với vai kia

Activator: chất kích hoạt là một chất hữu cơ có tính chất điều tiết, nó gắn với một nguyên tố có tính chất điều khiển và hoạt động tích cực, kích thích sự chuyển mã của gen cấu trúc

Adenosine triphosphate (ATP): là một nucleoside triphosphate cung cấp nguồn năng lượng quan trọng cho quá trình biến dưỡng. Nó được dự trữ dưới dạng phân tử trong tế bào

Allele: alen là một trong nhiều dạng của một gen định vị trên một locus của một nhiễm thể

Allelic complementation: là hiện tượng sản xuất ra một kiểu hình hoang dại hoặc gần như hoang dại của sinh vật mang hai alen đột biến khác nhau

Allelic exclusion: dùng để mô tả sự thể hiện trong tế bào plasma của các gen kháng thể được tìm thấy có một phần thuộc về một cặp nhiễm thể tương đồng

Allopolyploidy: dị đa bội là một đa bội mà trong đó các bộ nhiễm thể dư có nguồn gốc từ việc lai khác loài hay khác nhóm

Allosteric control: có quan hệ đến khả năng tương tác tại một đoạn của một protein, gây ảnh hưởng đến hoạt động của một đoạn khác

Allozyme: là dạng phân tử khác nhau của một enzyme được mã hóa bằng các dạng allelic khác nhau của một gen ở tại một locus nào đó

Alu family: là một nhóm phổ biến nhất của các đoạn DNA mã hóa có tính chất phân tán và có liên quan trong genome của con người, mỗi đoạn có qui mô #300bp, và được sao lập lại chừng 500 nghìn lần

Amber codon: là codon kết thúc của chuỗi, codon UAG trong mRNA, trong sinh tổng hợp protein

Amber mutation: dùng để mô tả bất cứ sự thay đổi nào trong DNA, tạo ra một amber codon ở tại điểm, mà nơi đó bị khống chế bởi một codon đặc trưng cho một amino acid của một protein

Amber suppressors: là các gen đột biến, mã hóa đối với tRNA, mà những anticodons của nó đã được kích hoạt để có thể nhận UAG codon cũng như các codon trước đó

Amplification: sự phóng đại thường được dùng để chỉ việc sản xuất các bản sao bổ sung của một đoạn nhiễm thể, được tìm thấy ở DNA bên trong hay bên ngoài nhiễm thể

Aneuploid: thể bội không chính dùng để chỉ nhiễm thể có cấu trúc khác với tình trạng lưỡng bội như thường dùng, do sự mất đi hay thêm vào các nhiễm thể hoặc một đoạn của nhiễm thể

Annealing: dùng để chỉ sự bắt cặp của dây DNA đơn với primer để hình thành thể dây xoắn đôi

Anticodon: là một chuỗi gồm 3 nucleotide của phân tử tRNA bắt cặp với codon tương ứng của mRNA, từ đó hình thành chuỗi polypeptide

Autosomes: là những nhiễm thể không phải nhiễm thể giới tính

B lymphocytes (B cells): tế bào bạch huyết là những tế bào có chức năng tổng hợp các kháng thể

Backcross: hồi giao hoặc lai thử nghiệm (test cross) là lai giữa F1 với bố hoặc mẹ

Bacteriophage: thực khuẩn thể là hiện tượng các virus xâm nhiễm vi khuẩn, còn được gọi tắt là phage

Base pair: nối base là cầu nối hydrogene giữa một purine và một pyrimidine, A với T, C với G

Bivalent: là sự kết hợp của hai nhiễm thể tương đồng vào lúc bắt đầu phân bào giảm nhiễm

Breakage and reunion: là hiện tượng mô tả cách thức tái tổ hợp gen, trong đó hai phân tử dây xoắn đôi DNA bị tách ra tùy theo vị trí và sau đó được nối lại

C banding: là kỹ thuật tạo ra các vùng nhuộm phẩm màu xung quanh tâm động

Cvalue: là tổng số các DNA của genome đơn bội

cDNA: complementary DNA, DNA được tổng hợp từ RNA thông qua men reverse transcriptase

cDNA clone: là một chuỗi DNA xoắn đôi, được mang nhờ một vectơ "cloning"

Centrioles: hạt trung tâm, trung tử ở đỉnh cực trong phân bào

Centromere: trung tâm đoạn, vùng gắn con thoi

Chiasma (pl. chiasmata): tần suất quần thể quần chéo, nơi quần chéo hai nhiễm thể tương đồng sẽ có hiện tượng trao đổi trong suốt quá trình giảm nhiễm.

Chromatid: thanh nhiễm sắc là một đoạn của nhiễm thể. Mỗi nhiễm thể có hai chromatid, nối với nhau tại tâm động

Chromatin: chất nhiễm sắc là vật liệu có trong nhiễm thể, bao gồm các DNA và protein

Chromomere: hạt nhiễm sắc là vùng có trong nhiễm thể, ăn màu, mang nhiều gen

Chromonema: (pl. chromonemata) sợi nhiễm sắc

Chromosome: nhiễm sắc thể, gọi tắt là nhiễm thể, đó là đơn vị của genome mang các gen, nơi dự trữ và di truyền các thông tin di truyền

Chromosome walking: dùng để diễn tả trạng thái cô lập các dòng mang các đoạn mật mã trùng lặp, chồng lên nhau, cho phép một vùng rộng lớn của nhiễm thể được kéo dài ra. "Walking" thường được hoàn thiện để tiến đến một locus đặc biệt nào đó có giá trị mà mình mong muốn

Chromosome alberation: sai hình nhiễm sắc thể là hiện tượng diễn tả một sự thay đổi nào đó về hình thái hoặc số lượng nhiễm thể

cis arrangement: là sự xếp đặt của một cá thể song dị hợp, trong đó có hai alen đại liên kết nhau trên một thể tương đồng, và hai alen đột biến lặn xảy ra ở một thể khác

cis-trans effect: là một ảnh hưởng có tính chất vị trí, tại đó có hai cá thể, ứng với từng mặt giống nhau về mặt di truyền, cho kết quả khác nhau về kiểu hình, giống như tổ chức các đoạn đột biến của nhiễm thể, thể cis cho kiểu hình bình thường hơn kiểu trans

Cistron: gen - đơn vị di truyền chức năng, là một đơn vị di truyền cis/trans test: tương đương với gen, hay một đoạn DNA được mã hóa đối với một polypeptide

Clone: dòng vô tính là một quần thể tế bào đồng nhất về di truyền

Cloning vector: là một plasmid hay phage được dùng để mang DNA lạ du nhập vào nhằm sản xuất nhiều hơn vật liệu hay sản phẩm protein

Codominant alleles: alen đồng trội dùng để diễn tả trạng thái dị hợp của cả hai alen.

Codomonant thường được dùng trong trường hợp mà hai sản phẩm có thể được khám phá trong thể dị hợp, cái này là thể allelic của cái kia

Codon: là một triplet (bộ ba) của những nucleotide, RNA hay DNA triplet, từ đó xác định một amino acid đặc biệt nào đó hay dấu hiệu cho thấy sự giải mã được kết thúc

Colony hybridization: là kỹ thuật lai *in-situ* để xác định vi khuẩn mang chimeric vectors mà DNA của vectơ này thì tương đồng với đoạn mã hóa đặc biệt nào đó

Compatibility group: nhóm tiếp hợp của những plasmid bao gồm các phân tử không thể cùng xuất hiện trong cùng một tế bào vi khuẩn

Complex locus: là nhóm các liên kết gen rất chặt, có quan hệ rất gần về mặt chức năng

Concatemer: của DNA gồm có một loạt các genome đơn vị được lặp lại một cách liên hoàn

Condensation reaction: phản ứng tập trung là phản ứng trong đó cầu nối được hình thành với sự mất đi một phân tử nước, thêm vào một amino acid trong quá trình hình thành chuỗi polypeptide

Conjugation: sự tiếp hợp dùng để mô tả hiện tượng giao phối giữa hai tế bào vi khuẩn, khi ấy một phần của nhiễm thể hoặc cả nhiễm thể của con này được chuyển sang con kia

Constitutive gene: gene cấu trúc thể hiện chức năng tương tác của RNA polymerase với promoter, không có điều chỉnh bổ sung, đôi khi nó còn được gọi là household genes trên khái niệm diễn tả chức năng đối với tất cả tế bào ở mức độ thấp

Contig (contiguous sequence): một đoạn dài sequence được hình thành từ một số các đoạn phân tử ngắn, chồng lấp nhau của sequence

Coordinate regulation: dùng để diễn đạt khái niệm kiểm soát tổng quát của một nhóm gen

Cordycepin: là 3' deoxyadenosine, một chất cản phản ứng polyadenylation của RNA

Core DNA: là DNA có qui mô 146bp ở trong một core particle

Core particle: là một sản phẩm của nucleosome bị phân cắt cho ra histone octamer, bao quanh nó là 146 nối base của DNA, còn được gọi là nucleosome bead

Corepressor: là một phân tử nhỏ làm kích hoạt sự kìm hãm gen trong khi giải mã bằng cách gắn chặt với một regulator protein

Cosmids: là plasmids, trong đó cos sites ở pha lambda được lồng vào, nhờ vậy plasmid DNA này có thể tập hợp *in vitro* trong pha "coat"

Crossing-over: trao đổi chéo là hiện tượng trao đổi vật liệu giữa hai nhiễm thể xảy ra trong phân bào giảm nhiễm và phù hợp với tái tổ hợp di truyền

Cruciform: là cấu trúc được tạo ra tại thời điểm tái tạo DNA, nếu cặp đôi chuỗi mã DNA nằm trên cùng một dây đơn

Cryptic satellite: là chuỗi DNA vệ tinh chưa được xác định theo gradient về trọng lượng riêng, mà nó tồn tại trong dây chính của DNA

ctDNA: DNA của thể lạp (chloroplast)

Cyclic AMP (cAMP): là phân tử AMP, trong đó gốc phosphate đính ở vị trí 3' và 5' của ribose, sự kết dính này làm kích hoạt CAP - một chất kích thích trong quá trình giải mã ở ngoài nhân

Cytoplasmic inheritance: di truyền tế bào chất dùng để diễn tả tính chất của gen định vị ở ty thể bộ hoặc thể lạp, hoặc ở cơ quan khác ngoài nhân

Cytoplasmic protein synthesis: sinh tổng hợp protein ở tế bào chất là sự giải mã của mRNA đại diện cho gen ở trong nhân, nó xảy ra tại ribô thể đính trên khung tế bào

Cytosol: dùng để diễn tả thể tích chung của tế bào chất, trong đó các cơ quan (như ty thể bộ) định vị

D loop: là vùng mà trong đó DNA của ty thể bộ có một đoạn ngắn RNA bắt cặp vào một đoạn dây đơn của DNA này, thay thế phần nguyên thủy dây đơn DNA trong vùng này.

Degeneracy: sự thoái hóa là thuật ngữ dùng để diễn tả mã di truyền có liên quan đến sự thiếu một ảnh hưởng nào đó đến base số ba của codon trong amino acid

Deletion: sự thiếu đoạn nhiễm sắc do sự lấy đi một chuỗi DNA, các vùng ở cùng một ph trở nên kết dính lại với nhau

Denaturation: sự biến chất của DNA hoặc RNA là sự đảo ngược từ dây đôi sang dây đơn do năng lượng nhiệt hoặc tác nhân hóa học làm gãy cầu nối hydrogen

Deoxy ribose nucleic acid (DNA): là phân tử của vật chất di truyền, nó gồm bốn base có nitrogen là adenine, guanine, cytosine, thymine, đính trên dây deoxy ribose và phosphate, tạo thành dây xoắn đôi hay còn gọi là duplex molecule

Derepressed state: là trạng thái của một gen được mở. Nó đồng nghĩa với chữ "induced" (kích thích) trong khi mô tả tình trạng thông thường của một gen, nó cũng đồng nghĩa với chữ "constitutive" trong khi mô tả ảnh hưởng của đột biến

Dimer: là một hỗn hợp được tạo ra giữa hai phân tử đồng nhất, do vậy có cùng một số hợp chất theo tỉ lệ phần trăm, nhưng trọng lượng phân tử thì gấp đôi so với phân tử nguyên thủy

DNA fingerprinting: là kỹ thuật dựa vào các DNA được phân cắt bằng các enzyme đặc hiệu

DNA polymerase: là enzyme xúc tác trong phản ứng sinh tổng hợp DNA hoặc DNA template của tiền chất deoxyribonucleotide triphosphate

Direct repeats: lặp lại trực tiếp là những chuỗi mã di truyền đồng nhất đại diện cho một hoặc hai bản sao theo cùng một hướng của một phân tử DNA, nó không cần thiết ở kế bên nhau

Disjunction: sự đứt đoạn mô tả hoạt động của các phân tử của một cặp nhiễm thể tại cực đối nghịch trong lúc phân bào

Dominant: tính trội của một alen so với một alen khác trong thể dị hợp

Downstream: thuật ngữ dùng để chỉ chuỗi mã di truyền được xác định có hoạt động theo hướng biểu hiện ra của một vùng mã hóa ở dưới dòng của codon đầu tiên, chẳng hạn. Hướng hoạt động của DNA từ vị trí 3' đến vị trí 5', dần dần đến vị trí đầu tiên của tiến trình giải mã

Electrophoresis: điện di là kỹ thuật cho chạy các phân tử trên điện trường

Endocytosis: là tiến trình mà thông qua đó, protein ở mặt ngoài của tế bào được tác động, chuyển nạp vào tế bào ở trong các lớp mạng tế bào chất

Endonuclease: là những men phân cắt các nối của nucleic acid, chuyên tính trên RNA, dây đơn DNA hoặc dây đôi DNA

Epigenetic changes: sự thay đổi biểu sinh (phân sinh) có ảnh hưởng đến kiểu hình mà không làm thay đổi kiểu gen

Episome: là một plasmid có thể tổng hợp được DNA vi khuẩn

Epistasis: tính át khuất gen mô tả một trạng thái, trong đó, sự thể hiện của một gen bị che khuất bởi ảnh hưởng có tính chất kiểu hình của một gen hay một nhóm gen khác

Euchromatin: chất nhiễm sắc điển hình

Excision (of phage): sự cắt bỏ phage của episome, hoặc chuỗi mã nào đó, dùng để chỉ một sự phóng thích từ nhiễm thể chủ như là một phân tử DNA có tính chất autonomous

Exocytosis: là tiến trình đưa protein từ tế bào ra môi trường bên ngoài thông qua các lớp màng plasma

Exon: là một đoạn của một gen mang mã di truyền, biểu thị trong RNA

Exonucleases: là những men phân cắt các nối của nucleotides từ điểm cuối của một chuỗi polynucleotide, nó chuyên tính ở cả hai vị trí 3' và 5' của DNA hoặc RNA

Expression vector: là một cloning vector giúp cho tiến trình gắn vào của một chuỗi mã ở tại một vị trí định sẵn, nó làm nhiệm vụ giải mã

Expressivity: độ biểu hiện tính trạng di truyền

F factor (fertility factor): yếu tố F là giới tính của vi khuẩn hoặc fertility plasmid

Facultative heterochromatin: dị nhiễm sắc ngẫu nhiên là thuật ngữ dùng để mô tả trạng thái bất động của chuỗi mã di truyền, có những bản sao tích cực, thí dụ nhiễm thể giới tính X của người

Filter hybridization: lai qua lọc được hoàn thiện bằng cách ủ ấm một DNA đã được mở thành dây đơn, bắt độn chúng trên tấm giấy lọc nitrocellulose, với một dung dịch RNA hoặc DNA có chất phóng xạ đánh dấu

Fingerprinting: là kỹ thuật mà trong đó protein được khổng chế bằng enzyme tiêu hóa tương ứng, kết quả các đoạn peptide sẽ được phân ra nhờ một phối hợp của sắc ký trên giấy, hoặc điện di trên gel nhờ các marker phân tử

5' end: đoạn cuối của chuỗi nucleic acid được kết thúc bằng một gốc phosphate tự do

5' to 3' growth: sự tổng hợp một chuỗi nucleic acid bằng cách kết hợp 5' (PO_3) của đoạn cuối một nucleotide, với 3' (OH) của đoạn cuối một nucleotide sau cùng

Fluidity: trạng thái lỏng là một đặc tính của màng, cho thấy khả năng của lipid di chuyển trong cùng một lớp

Footprinting: là một kỹ thuật dùng để xác định vị trí trên DNA được gắn bởi những protein, bằng cách đó làm tăng cường việc bảo vệ của các cầu nối tại vùng này chống lại các men nuclease.

Forward mutation: là đột biến do sự hình thành của một mutant gene từ alen đại

Founder effect: là ảnh hưởng có liên quan đến sự xuất hiện của quần thể mà nhiều cá thể có cùng nhiễm thể (hoặc vùng của nhiễm thể) có nguồn gốc từ một tổ tiên

G banding: là kỹ thuật làm phát sinh một phần trong nhiễm thể ở giai đoạn metaphase, làm phân biệt rõ các phần tử của một nhóm đơn bội (bằng phẩm màu Giemsa)

Gamete: giao tử (tinh trùng hay trứng), có tính chất đơn bội thể

Gene: là đơn vị di truyền định vị tại locus của một nhiễm thể, có một hoặc nhiều ảnh hưởng chuyên biệt trong một sinh vật, nó có thể phối hợp với gen khác và tạo đột biến một cách độc lập. Nó có khả năng mang thông tin bằng mã hiệu trên vật liệu như polypeptide, rRNA, tRNA

Gene conversion: sự hoán chuyển gen là sự thay đổi một dây đơn của dây xoắn đôi DNA để giúp nó bổ sung bằng dây khác ở bất cứ vị trí nào mà nó đã bị mất cặp nối base

Gene flow: là sự thay đổi một dạng mới của alen do các migrants giữa hai quần thể của một loài

Genetic code: mã di truyền là các triplets (codon) của RNA và DNA mang những thông tin di truyền thông qua sự phối hợp của 20 amino acid trong tổng hợp polypeptide, từ lúc bắt đầu cho đến khi có dấu hiệu kết thúc công tác giải mã ở ribosome

Genetic engineering: công nghệ di truyền bao gồm cả phân tích genome, chuyển nạp gen, tạo dòng nhân bản vô tính,.. phát triển trên thành tựu của sinh học phân tử

Genetic marker: thể đánh dấu di truyền như RFLP, còn gọi là dấu chuẩn phân tử

Genome: là nội dung di truyền của một bộ phận nhiễm thể đơn, hoặc một bộ đơn của gen vi khuẩn hay virus

Genotype: kiểu gen là cấu trúc di truyền của một tế bào hay một cá thể

Guanine: là một base "purine" của DNA và RNA

Haploid: thể đơn bội (n)

Haplotype: kiểu đơn là một kết hợp đặc biệt của alen trong một vùng nào đó của nhiễm thể, có liên kết rất chặt với marker mà marker này có xu hướng truyền sang một đơn vị ở thế hệ sau, nó có liên quan đến một phần RFLP của DNA trong sơ đồ biểu đồ bức xạ

Hapten: là một phân tử nhỏ hoạt động như một antigene khi nó kết hợp với một protein

Hemizygote: bán hợp tử là một cá thể lưỡng bội, bị mất đi bản sao của một gen đặc biệt nào đó (thí dụ do bị mất một nhiễm thể), do đó nó chỉ còn một copy đơn

Heterochromatin: dị nhiễm sắc tố

Heterogametic sex: có cấu trúc nhiễm thể lưỡng bội $2A + XY$

Heterogenous nuclear (hn) RNA: bao gồm việc giải mã gen ở trong nhân do men RNA polymerase II, nó có một phân bố rộng về kích thước và tính ổn định thấp

Hetero multimeric protein: gồm có các tiểu đơn vị không đồng nhất, được mã hóa bằng nhiều gen khác nhau

Heterozygote: dị hợp tử là cá thể có hai alen khác nhau ở trên cùng một locus

Highly repeatitive DNA: là hợp phần đầu tiên dùng để tái kết hợp và nó được cân bằng với DNA vệ tinh

Histones: là những protein căn bản có trọng lượng phân tử thấp, DNA ở trong eukaryotes, tạo ra nucleosome, đơn vị căn bản của sợi nhiễm sắc

Homeotic genes: gen phát triển đồng dạng là gen thông qua đột biến bị chuyển đổi chức năng điều khiển, thí dụ chân côn trùng có thể được thay bằng râu

Homogametic sex: có cấu trúc nhiễm thể lưỡng bội là $2A + XX$

Homologues: nhiễm thể tương đồng mang các loci giống nhau, do đó một tế bào lưỡng bội sẽ có hai copy của mỗi homologue, có từ bố mẹ

Homomultimeric protein: gồm các tiểu đơn vị đồng nhất

Homozygote: đồng hợp tử là cá thể có cùng alen ở trên những loci tương ứng của những nhiễm thể tương đồng

Hyperchromicity: hiện tượng tăng sắc tố trên nhiễm thể là sự gia tăng mật số quang học khi DNA biến chất trở thành dây đơn

In-situ hybridization: sự lai RNA-DNA xảy ra trong quá trình DNA biến chất trở thành dây đơn

in vitro: xảy ra trong ống nghiệm, hay trong điều kiện môi trường nhân tạo

in-vivo: xảy ra trong cơ thể sinh vật

In patibility: không tiếp hợp là hiện tượng không có khả năng của plasmid cùng hiện hữu trong một tế bào. Nó là nguyên nhân của tính miễn dịch của plasmid

Intron: là một đoạn trong DNA, nó không mang mật mã di truyền như exon

Inversion: sự đảo đoạn là sự thay đổi của nhiễm thể trong đó có một đoạn bị xoay ngược lại 180° và được gắn vào như cũ

Inverted repeats: hai đoạn DNA đồng nhất, trên hai hướng đối nghịch nhau ở cùng một phân tử. Sự lặp lại như vậy sẽ tạo ra hiện tượng palindrome ($5'-3'$)

Isozymes: là hai hoặc nhiều enzyme có khả năng xúc tác trong cùng một phản ứng, nhưng nó có khác nhau một chút về cấu trúc và hiệu quả dưới những điều kiện ngoại cảnh khác nhau (thuật ngữ **allozymes** cũng có ý nghĩa tương tự)

Karyotype: là nền tảng nhiễm thể của một tế bào hay một cá thể, là thể hoàn toàn của nhiễm thể thường được quan sát lúc phân bào

Kb: chữ viết tắt của 1000 base pair (bp) của DNA hay RNA

Kinetic plexity: là phức hợp của một hợp phần DNA nào đó được đo bằng cơ học của sự tái tổ hợp DNA

Kinetochores: đoạn trung tâm, vùng gắn thoi

Lagging chain: dây đơn của DNA được tập trung một cách không liên tục tại thời điểm nó sao lập lại, ở vị trí $3'$ và $5'$

Lampbrush chromosomes: là những nhiễm thể rất to tìm thấy trong trứng của loài lưỡng thể (amphibia)

Leader: là đoạn mã hóa chưa được giải ở điểm cuối $5'$ của mRNA, có trước codon đầu tiên

Leader sequence: của một protein là một đoạn ngắn có N ở cuối, đáp ứng với việc đi vào, hoặc đi qua một màng tế bào

Lethal locus: là gen mà ở trong đó sự đột biến gây chết được ghi nhận

Library: thư viện, kho lưu trữ các dòng là một bộ các đoạn DNA, biểu thị cho toàn bộ genome

Ligation: sự buộc thắt là kiểu hình thành của một cầu nối phosphor diester giữa hai base ở kề bên nhau, được phân biệt tại dây đơn DNA

Linkage: sự liên kết diễn tả xu hướng của các gen gắn kết với nhau trong điều kiện tính

trạng di truyền nào đó, trên cùng một nhiễm thể, được đo bằng tần suất tái tổ hợp của các loci (centi Morgan)

Linkage group: bao gồm tất cả các loci có thể được kết hợp lại một cách trực tiếp hoặc gián tiếp bằng mối quan hệ liên kết, tương ứng với một nhiễm thể

Linkage disequilibrium: sự không cân bằng liên kết dùng để diễn tả một trạng thái mà trong đó, các tổ hợp của những marker di truyền xuất hiện với tần suất nhiều hơn hoặc ít hơn trong quần thể. Nó xác định nhóm marker nào có tính di truyền thật sự. Kết quả của sự tái tổ hợp bị giảm đi ở vùng mà trong đó nó không đủ thời gian để đạt sự cân bằng, vì một trong những marker này đã du nhập vào quần thể.

Linker DNA: là DNA nhân tạo được thêm vào DNA mục tiêu để vị trí ghi nhận (recognition site) của phân tử đáp ứng với một enzyme phân cắt nào đó, hay enzyme kết nối

Locus: vị trí đặc biệt được một gen nào đó định vị, trong một nhiễm thể

LOD score: là phương tiện đo liên kết gen, theo log thập phân. Ngưỡng thông thường có LOD score 3.0, có nghĩa là tỉ lệ 1000:1, so sánh với 50:1 - xác suất của một cặp loci bất kỳ, không có liên kết

LTR: viết tắt của chữ long-terminated repeat, là một đoạn mã được lặp lại trực tiếp ở cả hai đầu của một DNA retroviral

Lysis: sự phân giải

Lysogenic repressor: là một protein ngăn cản thực khuẩn thể non (prophage) phân giải

Lysogeny: là trạng thái mà trong đó vật liệu di truyền của một virus và ký chủ của nó là vi khuẩn được hợp nhất

Map unit: đơn vị bản đồ dùng để diễn tả khoảng cách di truyền giữa hai alen liên kết nhau, tương ứng với tần suất tái tổ hợp 1%. Map distance được đo bằng đơn vị cM (centi Morgan) = phần trăm tái tổ hợp

Marker (DNA): là một đoạn có kích thước được biết dùng trong điện di

Meiosis: gián phân giảm nhiễm

Melting temperature (T_m): là giá trị trung bình của một quãng nhiệt độ, nơi đó DNA bị biến chất, mở dây xoắn đôi thành dây đơn

Mitosis: gián phân đẳng nhiễm, sự phân bào nguyên nhiễm

mtDNA: là DNA của ty thể bộ

Multicopy plasmids: là những plasmid hiện diện trong vi khuẩn với số lượng lớn gấp nhiều lần plasmid có trong nhiễm thể

Multimeric proteins: là những protein có hơn một đơn vị phụ (subunit)

Multagenes: là chất làm tăng mức độ đột biến do sự thay đổi trong DNA

Myeloma: là một dòng của tế bào khối u có từ một tế bào bạch huyết, thường sản sinh một dạng đơn của immunoglobulin

Negative regulators: các chất điều tiết có nhiệm vụ ngắt phản ứng chuyển mã hay giải mã

Nick translation: mô tả khả năng của *E. coli* DNA polymerase I có thể sử dụng ở một thời điểm đúng như vị trí khởi đầu từ một dây đơn hay một dây xoắn kép DNA, làm cho dây này bị biến hóa và được thay thế bằng sự tổng hợp trở lại cá vật liệu mới. Phương pháp này thường được áp dụng trong du nhập các nucleotide có đánh dấu bằng đồng vị phóng xạ trong DNA *in vitro*.

Nonsense codon: là codon cuối cùng - bất cứ một trong ba triplet (UAG, UAA, UGA) gây ra sự kết thúc sinh tổng hợp protein. (UAG được xem như một amber, và UAA là orche)

Northern blotting: là kỹ thuật chuyển RNA từ agarose gel đến màng lọc nitrocellulose, tại đó, người ta có thể lai RNA với nhau để cho ra DNA

Nucleolar organizer: là vùng của một nhiễm thể mang gen mã hóa đối với rRNA

Nucleoid: là một thể compact có chứa genome trong một vi khuẩn

Nucleolus: là một vùng đặc biệt của nucleus, được tạo ra do sự giải mã của những gen rRNA

Nucleosome: là tiểu đơn vị căn bản có tính chất cấu trúc của chromatin bao gồm khoảng 200bp DNA và một octamer của histone

Nucleolytic: là phản ứng thủy phân một cầu nối phosphodiester trong một nucleic acid

Nucleoside: là một phân tử có pentose gắn với một base có nitrogen

Nucleotide: là đơn vị của nucleic acid bao gồm có một pentose gắn với một nhóm phosphate và một base có nitrogen

Orche codon: chuỗi kết thúc bằng codon UAA

Oncogene: một gen có thể bắt đầu và duy trì một trạng thái khối u trong một cơ thể sinh vật và điều này tạo ra hiện tượng proto oncogene của một tế bào bình thường

Opal codon: chuỗi kết thúc bằng codon UGA

Operator: đoạn DNA kết hợp với một hay nhiều gen cấu trúc, sự tương tác tạo ra một gen có tính chất điều hòa (regulatory) khi kiểm soát sự giải mã

Operon: một đơn vị của sự giải mã gồm có một hay nhiều gen cấu trúc kết hợp với một operator và một promoter

ORF (open reading frame): một vùng của sequence giữa những stop codon

Organelles: là những phần ở trong tế bào chất được màng tế bào bao bọc

Origin (ori): là một đoạn mã hóa của DNA, tại đó sự tự tái bản được bắt đầu

Orphans: được xác định trong gen có những vị trí đặc biệt, vị trí này được phân lập, nó có liên quan đến các phân tử của một gene cluster

Packing ratio: là tỉ lệ chiều dài của DNA so với chiều dài đơn vị của sợi (fiber) mang DNA ấy

Pairing of chromosome: xem synapsis

Palindrome: là một đoạn mã của DNA được đọc từ trái sang phải (5'-3'). Nó bao gồm các lặp lại kết nối với nhau

pPR322: là một trong các plasmid đầu tiên được sử dụng trong sinh học phân tử

PCR (polymerase chain reaction): mô tả kỹ thuật mà trong đó một chu kỳ bao gồm sự thoái hóa (denaturation), tác động lại ở đầu dây DNA (annealing) với những primer, và phát triển DNA nhờ polymerase. Kỹ thuật này dùng để khuếch đại số copy của một đoạn DNA nhiều gấp 10^6 lần

Periodicity of DNA: là số lượng cặp nối base trên một "turn" của dây xoắn đôi

Phage (bacteriophage): là một virus với một tế bào vi khuẩn - ký chủ của virus ấy

Phenotype: kiểu hình là sản phẩm của vật chất di truyền và môi trường

Plasmid: là một DNA có cấu trúc vòng, ở ngoài nhiễm thể, có khả năng tự sao chép

Pleiotropic gene: gen đa tính trạng, một gen điều khiển nhiều tính trạng

Polymorphism: thể đa hình nhằm mô tả sự có mặt đồng thời của quần thể trong genome biểu hiện biến dị có tính chất allelic, có thể quan sát trên alleles tạo ra các kiểu hình khác nhau, hoặc sự thay đổi của DNA ảnh hưởng đến phân giới hạn (restriction patterns)

Primer: là một đoạn mã DNA ngắn, thường dùng làm mồi để bắt cặp với một dây đơn DNA, và cho một gốc 3' OH tự do, nơi đó một DNA polymerase bắt đầu tổng hợp chuỗi deoxyribonucleotide

Prokaryotic organism: thí dụ vi khuẩn thiếu nuclei

Promoter: là một vùng DNA gồm có sự gắn kết của RNA polymerase để bắt đầu giải mã. Đoạn DNA chứa các ký hiệu bắt đầu có thể được ghi nhận bởi RNA polymerase khi bắt đầu giải mã

-10 sequence: là đoạn mã hóa có tác động liên ứng TATAATG tập trung khoảng 10 trước điểm khởi hành của một gen của vi khuẩn. Nó bao gồm nhiệt độ ban đầu melting DNA do RNA polymerase

-35 sequence: là đoạn mã hóa tập trung khoảng 35bp trước điểm khởi hành của một gen của vi khuẩn. Nó bao gồm sự ghi nhận đầu tiên của RNA polymerase

Puff: là sự phát triển của một band của một nhiễm thể polytene kết hợp với sự tổng hợp RNA ở cùng một locus trong band đó

Q band: là vùng ăn màu fluorescent dọc theo chiều dài nhiễm thể

Quaternary structure: cấu trúc protein trong đó có sự thể hiện kiến trúc theo kiểu multimeric

Quick stop: của DNA mutants của *E. coli* làm dừng lại sự sao chép ngay lập tức, khi nhiệt độ được tăng lên 42°C

Reading frame: chuỗi mã codon được xác định bởi sự đọc một sequence từ một điểm cho sẵn. Vì mật mã di truyền dựa trên triplet của nucleotide, cho nên có 3 reading frame có thể có trên mỗi dây đơn

Recessive: alen lặn bị che khuất bởi alen trội trong kiểu hình dị hợp tử

Recombinant: con lai tái tổ hợp có kiểu gen khác nhau từ một nguồn bố mẹ

Recombinant DNA: là phân tử DNA lai với một pr e có đánh dấu bằng đồng vị phóng xạ

Recombinant joint: là vị trí mà nơi đó hai phân tử DNA xoắn kép phối hợp nhau

Regulatory gene: gen mã hóa đối với một RNA hoặc protein có chức năng kiểm soát sự thể hiện của các gen khác

Renaturation: sự hoàn nguyên là hiện tượng tái phối hợp lại hai dây đơn DNA để nó trở thành dây xoắn kép

Repetition frequency: là số lần sao chép của một đoạn mã hóa nào đó có trong genome đơn bội, tương đương với 1 trong trường hợp nonrepetitive DNA, và > 2 trong trường hợp repetitive DNA

Repetitive DNA: DNA sao chép lại xảy ra trong phản ứng tái kết hợp, mặc dù có nhiều đoạn mã trong hợp chất đó, nó vẫn cho phép bất cứ một cặp DNA mã hóa nào có tính chất phụ trợ (plementary) đều có thể tái kết hợp (reassociate)

Reporter gene: là một đơn vị mã hóa mà sản phẩm của nó được trắc nghiệm một cách dễ dàng (thí dụ như chloramphenicol transacetylase); nó có thể được gắn với bất cứ một promoter nào sao cho sự thể hiện của gen này được dùng để thử nghiệm chức năng của promoter

Restriction enzyme: xác định một đoạn ngắn có mã hóa của DNA và cắt chúng ra tại những điểm thích ứng

RFLP (restriction fragment length polymorphism): đề cập đến tính khác nhau về di truyền ở tại các điểm tương ứng với các restriction enzyme, tạo ra sự khác nhau về chiều dài của đoạn DNA. Nó được dùng trong công tác lập bản đồ liên kết gen

Restriction map: là một sắp xếp theo đường thẳng các vị trí trên DNA được phân cắt bởi nhiều restriction enzyme khác nhau

Retroposon: là một transposon di động qua một RNA; nguyên tố DNA sẽ được giải mã qua RNA, rồi sau đó đi ngược trở lại DNA, để nó gắn vào một vị trí mới trong genome

Retroregulation: mô tả khả năng của một đoạn mã hóa downstream (gốc) điều chỉnh sự giải mã của một mRNA

Retrovirus: là một RNA virus phát triển nhờ sự thay đổi ngược lại thành dây xoắn đôi DNA

Reverse transcription: sự chuyển mã ngược là sự tổng hợp DNA trên một template của RNA, được hoàn thành nhờ reverse transcriptase enzyme

Reverse translation: sự giải mã ngược là kỹ thuật phân lập các gen (đối với RNA) nhờ khả năng của chúng trong sự lai với một đoạn mã oligonucleotide nào đó, đoạn này được chuẩn bị bằng các dự đoán đoạn mã nucleic acid từ những đoạn mã hóa của protein biết trước

Rho factor: là một protein gồm có assisting *E. coli* RNA polymerase để kết thúc phản ứng giải mã ở bất cứ vị trí nào (rho dependent)

Rho independent terminators: là những đoạn mã hóa của DNA làm cho *E. coli* RNA polymerase kết thúc *in vitro* với sự vắng mặt của rho factor

Rifamycins (bao gồm rifampicins): ngăn cản sự giải mã trong vi khuẩn

S phase: là restricted part của chu trình tế bào eukaryotic, chu trình này xảy ra trong sinh tổng hợp DNA

Saltatory replication: là một sự khuếch đại bất ngờ để sản sinh ra một số lượng lớn các bản sao của một đoạn mã hóa nào đó

Satellite DNA: gồm có nhiều đoạn ngắn phân tử lặp lại được gọi là basic repeating

Segmentation gene: là khái niệm đề cập đến sự kiểm soát số lượng các "body segments" trong côn trùng

Sex plasmid: thường là một episome, nó có thể khởi động tiến trình conjugation (tiếp hợp), nhờ đó, các vật liệu trong nhiễm thể của một vi khuẩn sẽ được chuyển đến cơ thể khác

Shuttle vector: là một plasmid được cấu tạo, có nguồn gốc đối với khả năng sao chụp cho hai ký chủ (thí dụ *E. coli* và *S. cerevisiae*) sao cho, người ta có thể dùng nó để mang các đoạn mã hóa lạ vào prokaryotype hoặc eukaryotype

Sigma factor: là subunit của bacterial RNA polymerase cần thiết cho sự khởi động, nó có ảnh hưởng chính trong sự chọn lọc ở những vị trí kết nối (promoters)

Single copy plasmids: được duy trì trong vi khuẩn ở một tỉ lệ là một plasmid cho mỗi nhiễm thể ký chủ

snRNA: viết tắt từ chữ "small nuclear RNA"

snRNPs: viết tắt từ chữ "small nuclear ribonucleoprotein"

SOS box: là đoạn mã hóa DNA (operator) của khoảng 20bp, được ghi nhận bằng LexA repressor protein

Southern blotting: mô tả qui trình chuyển DNA bị biến chất (dây đơn) từ agarose gel đến màng lọc nitrocellulose, nơi đó nó có thể được lai với một nucleic acid thăm dò

Spheroplast: là tế bào của vi khuẩn hay nấm men có thành tế bào bị lấy ra hoàn toàn hoặc một phần lớn

Splicing: là hành động mô tả sự lấy ra của intron và sự gắn vào của exon trong RNA, theo đó intron được "sliced out", còn exon được "sliced together"

Suppressor: (extragenic) thường là một gen mã hóa một tDNA đột biến có trên codon đột biến, theo ý nghĩa codon đầu tiên

Suppressor: (intragenic) là đột biến có tính chất bù đắp làm cho phục hồi khung đọc mã hiệu đầu tiên, sau khi có sự chuyển dịch khung

Synapsis: sự tiếp hợp là hiện tượng mô tả sự kết hợp của hai cặp "sister chromatids" tương trưng cho các nhiễm thể tương đồng xảy ra ở những giai đoạn đầu tiên của gián phân giảm nhiễm, kết quả nó tạo ra một cấu trúc được gọi là bivalent

T cells: là những bạch huyết cầu của T (thymic) lineage; nó có thể được phân ra làm nhiều loại chức năng khác nhau. Nó mang TcR (T cell receptor) và bao gồm phản ứng miễn dịch trong môi trường tế bào

Tail primer: là một primer bổ sung trên dây template ở vị trí 3', nhưng lại chứa thêm sequence khác, như một sequence có tính chất universal primer ở vị trí 5'.

Tm: viết tắt từ chữ "melting temperature" - nhiệt độ để DNA biến chất thành dây đơn

Tandem repeats: là những bản sao của cùng một đoạn mã hoá

Termination codon: là một trong ba triplet sequence, UAG (amber), UAA (ochre), hay UGA gây ra sự kết thúc sinh tổng hợp protein, nó còn được gọi là non-sense codon.

Terminator: là một đoạn mã hóa DNA, thể hiện ở cuối giai đoạn giải mã, gây ra RNA polymerase để kết thúc sự giải mã

Topoisomerase: là một enzyme, nó có thể thay đổi số liên kết của DNA (trong các giai đoạn 1 của loại hình số 1, giai đoạn 2 của loại hình II)

Transcription: sự chuyển mã là sự tổng hợp RNA trên một DNA template

Transcription unit: là khoảng cách giữa các điểm khởi đầu và kết thúc do hoạt động của RNA polymerase, nó có thể có nhiều hơn một gen

Transduction: sự truyền tính trạng là quá trình chuyển một gen của vi khuẩn đến một vi khuẩn khác thông qua phage của nó

Transfection: của những tế bào eukaryotic là sự có được các marker di truyền mới do sự đưa vào những DNA bổ sung

Transformation: sự chuyển nạp gen là hiện tượng đưa vào cơ thể sinh vật những gen mới

Translation: sự giải mã là hiện tượng tổng hợp protein trên mRNA template

Translocation: sự chuyển vị được dùng trong qui mô của nhiễm sắc thể

Transposase: là hoạt động của enzyme bao gồm sự gắn vào transposon ở một vị trí mới

Transposon: là một đoạn mã hóa DNA có thể tự nó gắn vào ở một vị trí mới trong genome (không cần một sự liên quan nào về mã di truyền tại target locus)

tRNA: viết tắt từ chữ "transfer RNA", RNA mang mã di truyền

Universal primer: là những primer tiêu chuẩn làm mẫu đối với các đoạn phân tử của M13 genome. Những sequence này đã và đang được tạo thành những vectơ phổ biến

Unscheduled DNA synthesis: là sinh tổng hợp DNA xảy ra bên ngoài phase S của tế bào eukaryotic.

Up promoter mutation: làm gia tăng tần suất của khởi động quá trình giải mã

Upstream: xác định đoạn mã hóa theo chiều hướng đối lập với downstream

URF: viết tắt từ chữ "open unidentified reading frame"

Virion: là lớp vỏ ngoài protein của một phân tử virus

Wilde type: những alen bình thường của một gen

Wobble hypothesis: diễn tả khả năng của một tRNA ghi nhận hơn một codon bằng các bắt cặp đặc biệt (không phải , A-T) ở base thứ ba của một codon

Writhing number: là số lần quấn chéo của một duplex axis trong không gian

DI TRUYỀN PHÂN TỬ

Chịu trách nhiệm xuất bản:

LÊ VĂN THỊNH.

Bản thảo: **Vũ Thị Hòa**

Sửa bài: **Bùi Chí Bửu, Nguyễn Thị Lang**

Maquette - Bìa: **Hồng Hà**

NHÀ XUẤT BẢN NÔNG NGHIỆP

D14 - Phương Mai - Đống Đa - Hà Nội

Điện thoại : (04) 8523887 - 8525070 - 8521940

CHI NHÁNH NHÀ XUẤT BẢN NÔNG NGHIỆP

58 Nguyễn Bình Khiêm, Q.1, TP. Hồ Chí Minh

Điện thoại : (08) 8297157 - 8299521

In 1020 bản khổ 19×27 cm tại Xí nghiệp in Quận I, TP.HCM.
Chấp nhận đề tài số 1354/XB-QLXB, ngày 31/12/1998.
In xong nộp lưu chiểu tháng 6/ 1999.

5-57.040
NN-99

192/1354-98

GIA : 40.000