



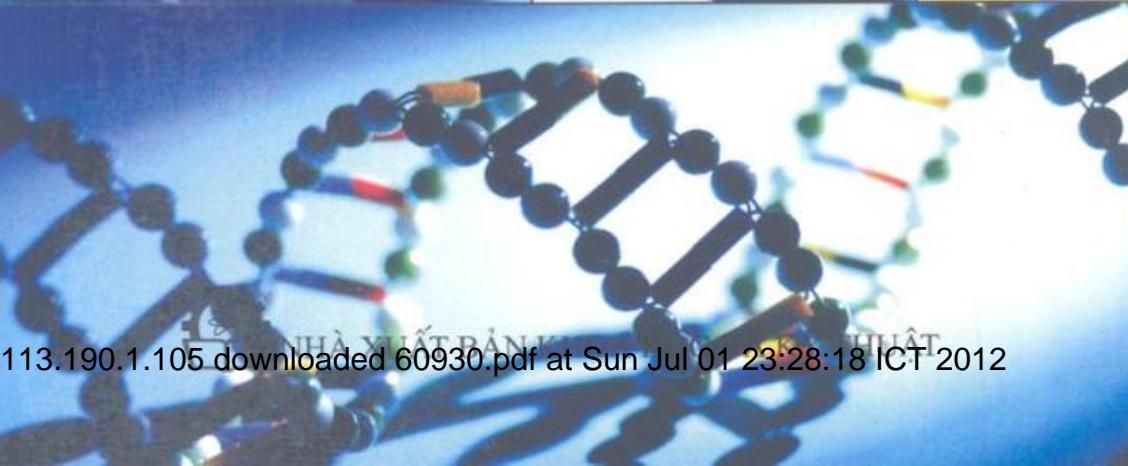
1956 - 2006

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  
50 NĂM XÂY DỰNG VÀ PHÁT TRIỂN

PGS. TS. KHUẤT HỮU THANH

# KỸ THUẬT GEN

## NGUYÊN LÝ VÀ ỨNG DỤNG





1956 - 2006

**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI  
50 NĂM XÂY DỰNG VÀ PHÁT TRIỂN**

PGS. TS. Khuất Hữu Thanh

# **KỸ THUẬT GEN nguyên lý và ứng dụng**



**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT  
HÀ NỘI**

## LỜI NÓI ĐẦU

*Kỹ thuật gen ngày càng có nhiều ứng dụng trong nông nghiệp và y học bảo vệ sức khoẻ con người. Nguyên lý kỹ thuật gen và các phương pháp nghiên cứu cơ bản trong công nghệ sinh học phân tử là những kiến thức cần thiết cho các nhà sinh học và kỹ sư công nghệ sinh học. Các thành tựu nghiên cứu bộ gen người, bộ gen cây lúa, và nhân bản động vật... đã mở ra một thời kỳ mới trong sinh học, giúp con người hiểu biết rõ bản chất của nhiều loại bệnh di truyền, bệnh truyền nhiễm, chẩn đoán bệnh sớm, góp phần điều trị có hiệu quả và bảo vệ sức khoẻ con người.*

*Đáp ứng nhu cầu học tập của sinh viên, chúng tôi biên soạn cuốn giáo trình **Kỹ thuật gen nguyên lý và ứng dụng**, nhằm mục đích cung cấp kiến thức cơ bản trong lĩnh vực di truyền phân tử, các phương pháp nghiên cứu cơ bản trong công nghệ sinh học phân tử và một số ứng dụng kỹ thuật gen trong thực tiễn.*

*Giáo trình được biên soạn trên cơ sở chương trình giảng dạy cải tiến, đáp ứng nhu cầu nâng cao chất lượng đào tạo kỹ sư công nghệ sinh học. Kiến thức trong giáo trình được cập nhật, bao gồm các kiến thức khoa học cơ bản và ứng dụng thực tiễn, giúp người đọc nắm vững cơ sở khoa học của các ứng dụng. Giáo trình được dùng làm tài liệu học tập hoặc tham khảo cho sinh viên, học viên cao học, nghiên cứu sinh... các ngành Công nghệ sinh học, Công nghệ thực phẩm Viện Công nghệ sinh học và Công nghệ thực phẩm, Trường Đại học Bách khoa Hà Nội, Viện Đại học mở Hà Nội và sinh viên ngành Sinh học các trường đại học khác...*

*Trong quá trình biên soạn giáo trình, không thể tránh khỏi những thiếu sót. Chúng tôi mong muốn nhận được nhiều ý kiến đóng góp của bạn đọc và đồng nghiệp để giáo trình ngày càng hoàn chỉnh hơn.*

*Xin chân thành cảm ơn các ý kiến đóng góp của bạn đọc.*

**PGS.TS. Khuất Hữu Thanh**

**MỤC LỤC**

	Trang
Lời nói đầu .....	3
Mục lục .....	4
Chữ viết tắt và kí hiệu.....	10

**PHẦN I****TỪ PHÂN TÍCH BỘ GEN ĐẾN DỤNG CÂY PHÂN LOẠI..... 13****Chương 1.**

<b>Đặc điểm cấu trúc bộ gen và gen của <i>prokaryote</i> và <i>eukaryote</i>.....</b>	<b>14</b>
I. Khái niệm bộ gen.....	14
II. Lược sử nghiên cứu bộ gen và giải trình tự bộ gen.....	15
III. Bộ gen một số nhóm sinh vật .....	16
1. Kích thước bộ gen và số lượng gen của một số đại diện điển hình.....	16
2. Cấu trúc bộ gen.....	17
2.1. Đặc điểm di truyền và cấu trúc bộ gen virus.....	17
2.2. Đặc điểm di truyền và cấu trúc bộ gen <i>prokaryote</i> .....	23
2.3. Đặc điểm di truyền và cấu trúc bộ gen <i>eukaryote</i> .....	28
IV. Cấu trúc gen của <i>prokaryote</i> và <i>eukaryote</i> .....	39
1. Cấu trúc gen mã hóa protein.....	39
2. Đặc điểm cấu trúc gen mã hóa protein của <i>prokaryote</i> .....	41
3. Đặc điểm cấu trúc gen mã hóa protein của <i>eukaryote</i> .....	43
Tài liệu tham khảo chính.....	46

**Chương 2.**

<b>Bản đồ di truyền.....</b>	<b>48</b>
I. Vài nét về lược sử nghiên cứu bản đồ di truyền.....	48
II. Một số loại bản đồ di truyền chủ yếu .....	50
1. Bản đồ hình thái nhiễm sắc thể.....	51
2. Bản đồ di truyền liên kết.....	53
3. Bản đồ di truyền giao nap.....	57
4. Bản đồ di truyền tế bào.....	59
5. Bản đồ lai phóng xạ .....	61
6. Bản đồ giới hạn.....	64
7. Bản đồ vật lí.....	65
8. Bản đồ trình tự gen .....	66
<i>Tài liệu tham khảo chính.....</i>	<i>68</i>

**Chương 3.**

<b>Phân tích dữ liệu di truyền và phương pháp dựng cây phân loại .....</b>	<b>69</b>
I. Ý nghĩa của phân tích dữ liệu di truyền và dựng cây phân loại .....	69
II. Các dạng cây phân loại.....	71
1. Cây phân loại không gốc .....	71
2. Cây phân loại có gốc.....	72
3. Số lượng và số dạng cây phân loại.....	73
III. Khoảng cách di truyền .....	74
1. Khái niệm khoảng cách di truyền.....	74
2. Phương pháp xác định khoảng cách di truyền với dữ liệu trình tự các nucleotid trong gen và bộ gen.....	75
3. Phương pháp xác định khoảng cách di truyền trên cơ sở các dấu chuẩn vi vệ tinh trong gen và bộ gen .....	78
4. Phương pháp xác định khoảng cách di truyền với dữ liệu trình tự các acid amino trong chuỗi polypeptid.....	80

5. Phương pháp xác định khoảng cách di truyền với dữ liệu tần số allele của các locus gen.....	80
<b>IV. Phương pháp dựng cây phân loại với dữ liệu khoảng cách di truyền .....</b>	<b>84</b>
1. Nguyên tắc dựng cây phân loại.....	84
2. Các bước dựng cây phân loại trên cơ sở khoảng cách di truyền của các trình tự DNA, RNA và protein.....	86
3. Phương pháp lập ma trận khoảng cách di truyền và dựng cây phân loại .....	87
4. Sử dụng các phần mềm trong phân tích số liệu và dựng cây phân loại .....	95
<i>Tài liệu tham khảo chính.....</i>	99

**PHẦN II****NGUYÊN LÝ VÀ PHƯƠNG PHÁP NGHIÊN CỨU CƠ BẢN****CỦA KỸ THUẬT GEN..... 101****Chương 4.**

<b>Nguyên lý kỹ thuật gen .....</b>	<b>102</b>
I. Một số yếu tố cần thiết trong kỹ thuật gen.....	102
1. Chuẩn bị gen cần tách dòng.....	102
2. Enzym giới hạn.....	108
3. Các loại enzym thường sử dụng trong kỹ thuật gen.....	112
4. Vectơ tách dòng.....	116
5. Vectơ biểu hiện gen .....	125
6. Các hệ thống biểu hiện gen.....	127
II. Nguyên lý kỹ thuật gen .....	130
1. Các bước chủ yếu của kỹ thuật gen .....	130
2. Nguyên tắc chọn vectơ tách dòng .....	132
3. Nguyên tắc tạo vectơ tái tổ hợp .....	133

4. Một số kỹ thuật biến nạp vectơ tái tổ hợp vào tế bào chủ.....	134
5. Kỹ thuật chọn lọc các dòng tế bào tái tổ hợp.....	136
<b>III. Các phương pháp tách dòng gen.....</b>	<b>140</b>
1. Tách dòng <i>in vitro</i> .....	140
2. Tách dòng <i>in silico</i> .....	144
<b>IV. Ngân hàng bộ gen .....</b>	<b>147</b>
1. Khái niệm ngân hàng bộ gen .....	147
2. Thiết lập ngân hàng bộ gen.....	148
<b>V. Ngân hàng cDNA .....</b>	<b>149</b>
1. Kỹ thuật phiên mã ngược tạo cDNA .....	150
2. Thiết lập ngân hàng cDNA.....	152
<i>Tài liệu tham khảo chính.....</i>	153

## Chương 5.

<b>Phương pháp PCR và các kỹ thuật cơ bản trong công nghệ sinh học phân tử .....</b>	<b>154</b>
<b>I. Phương pháp PCR (Polymerase Chain Reaction).....</b>	<b>154</b>
1. Nguyên lý chung của phương pháp PCR .....	154.
2. Các yếu tố cần thiết trong phản ứng PCR .....	156
3. Các ứng dụng chủ yếu của PCR.....	160
4. Một số kỹ thuật PCR chủ yếu.....	160
<b>II. Phương pháp điện di .....</b>	<b>170</b>
1. Nguyên lý chung .....	170
2. Các kỹ thuật điện di chủ yếu .....	171
3. Các bước thực hiện kỹ thuật điện di .....	172
<b>III. Phương pháp giải trình tự gen.....</b>	<b>173</b>
1. Giải trình tự gen theo phương pháp Maxam và Gilbert .....	173
2. Giải trình tự gen theo phương pháp dideoxy.....	176
3. Giải trình tự gen bằng máy tự động (Sequencer) .....	178
<b>IV. Các kỹ thuật lai phân tử.....</b>	<b>180</b>

1. Kỹ thuật lai Southern blotting.....	180
2. Kỹ thuật lai Northern blotting .....	182
3. Kỹ thuật lai Western blotting .....	183
4. Kỹ thuật lai tại chỗ .....	183
<b>V. Một số kỹ thuật phân tử sử dụng trong xác định tính đa dạng</b>	
di truyền của sinh vật và ứng dụng trong phân loại phân tử.....	184
1. Kỹ thuật RAPD.....	184
2. Kỹ thuật RFLP.....	186
3. Kỹ thuật AFLP .....	188
4. Kỹ thuật VNTR .....	189
5. Kỹ thuật DNA microarray .....	191
<b>VI. Các phương pháp chuyển gen.....</b>	<b>194</b>
1. Chuyển gen bằng thiết bị siêu âm.....	195
2. Chuyển gen bằng kỹ thuật điện xung (Electroporation) .....	196
3. Chuyển gen bằng kỹ thuật PEG.....	197
4. Chuyển gen bằng kỹ thuật vi tiêm .....	198
5. Chuyển gen bằng súng bắn gen .....	199
6. Chuyển gen bằng kỹ thuật sốc nhiệt.....	201
7. Kỹ thuật chuyển gen qua ống phún .....	202
8. Chuyển gen nhờ vi khuẩn <i>Agrobacterium tumefaciens</i> .....	203
<i>Tài liệu tham khảo chính.....</i>	205

**PHẦN III****MỘT SỐ ỨNG DỤNG CHỦ YẾU CỦA KỸ THUẬT GEN..... 207****Chương 6.**

<b>Ứng dụng kỹ thuật gen trong y học và chẩn đoán.....</b>	<b>208</b>
<b>I. Kỹ thuật gen trong sản xuất các chế phẩm sinh học .....</b>	<b>209</b>
1. Sản xuất insulin tái tổ hợp .....	209

2. Sản xuất hormon sinh trưởng người tái tổ hợp.....	216
3. Sản xuất interferon tái tổ hợp.....	221
4. Sản xuất yếu tố VIII và yếu tố IX tái tổ hợp.....	229
5. Sản xuất vaccin tái tổ hợp và vaccin DNA.....	233
II. Sử dụng kỹ thuật gen trong chẩn đoán và chữa bệnh.....	243
1. Sử dụng kỹ thuật gen trong chẩn đoán và điều trị ung thư do virus.....	243
2. Chẩn đoán một số bệnh nhiễm trùng bằng kỹ thuật PCR .....	245
3. Sử dụng kỹ thuật gen trong chẩn đoán và điều trị bệnh máu khó đông ( <i>Hemophilia</i> ) ở người.....	248
4. Sử dụng kỹ thuật microarray trong chẩn đoán bệnh .....	249
5. Sử dụng kỹ thuật gen trong chẩn đoán bệnh nhiễm trùng ở vật nuôi .....	251
6. Sử dụng kỹ thuật gen trong chẩn đoán bệnh cây trồng và kiểm tra thực phẩm chuyển gen.....	252
<i>Tài liệu tham khảo chính.....</i>	253

## Chương 7.

Kỹ thuật gen ứng dụng trong thực tiễn sản xuất.....	255
I. Kỹ thuật gen và các ứng dụng trong sản xuất nông nghiệp.....	255
1. Cây trồng biến đổi gen .....	255
2. Động vật biến đổi gen và động vật nhân bản vô tính.....	259
3. Kỹ thuật gen và thuốc trừ sâu sinh học.....	266
II. An toàn của sinh vật biến đổi gen.....	267
<i>Tài liệu tham khảo chính.....</i>	269

**CHỮ VIẾT TẮT VÀ KÍ HIỆU**

A	Nucleotid Adenin (Adenine)
a.a	acid amin
ADN	Axit deoxyribonucleic (Deoxyribonucleic Acid)
AFLP	Đa hình di truyền độ dài các đoạn DNA khuếch đại (Amplified Fragment Length Polymorphism)
bp	Cặp bazơ nitơ (Base pairs)
CHCl <sub>3</sub> - IsoA	Chloroform: isoamylalcohol (24: 1)
CHO	Tế bào trứng chuột đất vàng Trung Quốc (Chinnes Hamster Ovary)
CTAB	Cetyl Trimethyl Amoni Bromit
C	Cytosine
cDNA	DNA bổ sung (Complementary DNA)
cM	centi Morgan
cR	centi ray
DDBJ	Ngân hàng Dữ liệu Nhật Bản (DNA Data Bank of Japan)
DNA	Deoxyribonucleic acid
dNTP	Deoxyonucleotit triphosphat
EDTA	Ethylen Diamin Tetra Axetic
EtBr	Ethidi Bromit (Ethidium Bromide)
EMBL	Ngân hàng Dữ liệu Châu Âu (European Molecular Bioinformatic Laboratory)
EST	Expressed Sequence Tag
FASTA	Fast alignment search tool
FISH	fluorescent in situ hybridization
G	Guanine
G-C	Guanine-Cytosine
IGH	Hormon sinh trưởng người (Growth Hormone)

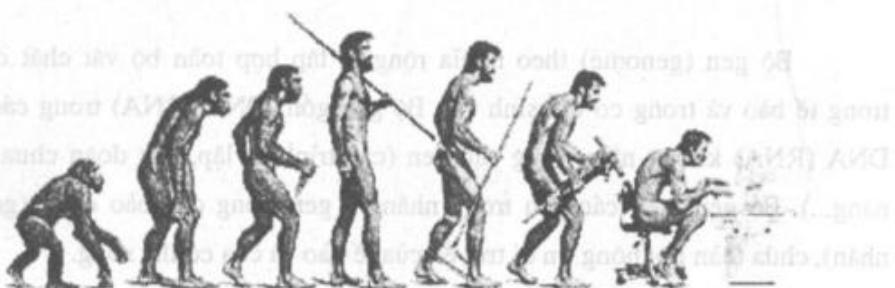
IFN	Interferon
kb	kilo bazơ (1 kb = 1000 bp)
kD	kilo Dalton
LB	Môi trường nuôi vi sinh vật Lucia Bertani
LINE	Trình tự 1 - 5 kb lặp lại dài (Long interspersed elements)
LTR	Đoạn lặp lại dài (Long Terminal Repeats)
Mb	Mega bazơ (1Mb = 1 000 000 bp)
ME	minimum evolution
mRNA	RNA thông tin (messenger RNA)
NJ	Neighbour Joining
NST	Nhiễm sắc thể
ORF	Khung đọc mã mở (Open Reading Frame)
PCR	Polymerase Chain Reaction
PDB	Ngân hàng Dữ liệu protein (Protein Databank)
PEG	Polyethylen Glycol
PHYLIP	Phần mềm PHYLIP (PhyLogeny Inference Package)
RAPD	Đa hình các đoạn ADN khuếch đại ngẫu nhiên (Random Amplify Polymorphism DNA)
RFLP	Đa hình các mảnh cắt giới hạn (Restriction Fragment Length Polymorphism)
RE	Enzym giới hạn (Restriction Enzyme)
RH	Bản đồ lai phóng xạ (radiation hybrid mapping)
RNA	Ribonucleic acid
rRNA	RNA riboxom (ribosomal RNA)
SDS	Natri Dodecyl sulfat (Sodium dodecyl sulphate)
SINE	Trình tự lặp lại ngắn 200 -300 bp (Short interspersed elements)
SSR	Đoạn lặp đơn giản (Simple seqences repeats)
STS	Trình tự đích (Sequence Target Site)
TAE	Tris - Axit axetic - EDTA

TE	Tris - EDTA
Tm	Nhiệt độ nóng chảy (Melting Temperatures)
Tn	Gen nhảy (Transposon)
T	Thymine, Thymidine
tRNA	RNA vận chuyển (transfer RNA)
U	Uracil
UPGMA	Unweighted Pair Group Method using arithmetic Averages
UTR	Vùng không dịch mã (Untranslated Region)
VNTR	Đoạn lặp lại (Variable Number of Tandem Repeat).

## Chương I

## PHẦN I

## TỪ PHÂN TÍCH BỘ GEN ĐẾN DỰNG CÂY PHÂN LOẠI

**Chương 1:****Đặc điểm cấu trúc bộ gen và gen của prokaryote và eukaryote****Chương 2:****Bản đồ di truyền và ứng dụng****Chương 3:****Phân tích dữ liệu di truyền và dựng cây phân loại**

## Chương 1

### **ĐẶC ĐIỂM CẤU TRÚC BỘ GEN VÀ GEN CỦA PROKARYOTE VÀ EUKARYOTE**

#### **I. Khái niệm bộ gen**

Bộ gen (genome) theo nghĩa rộng là tập hợp toàn bộ vật chất di truyền trong tế bào và trong cơ thể sinh vật. Bộ gen gồm DNA (RNA) trong các gen và DNA (RNA) không nằm trong các gen (các trình tự lặp, các đoạn chưa rõ chức năng...). Bộ gen gồm các gen trong nhân và gen trong các bào quan (gen ngoài nhân), chứa toàn bộ thông tin di truyền của tế bào và của cơ thể sống.

Các loài, giống, cá thể sinh vật khác nhau có bộ gen đặc trưng riêng về cấu trúc vật chất di truyền, hàm lượng DNA, số lượng gen, đặc điểm phân bố của các nhóm DNA và trình tự sắp xếp của các nucleotid trong DNA (RNA).

Virus chưa có cấu tạo tế bào, bộ gen virus có cấu trúc tương đối đơn giản, số lượng gen ít. Các họ virus khác nhau có cấu trúc vật chất di truyền khác nhau mang tính đặc trưng riêng. Vật chất di truyền của các virus khác nhau có thể là RNA mạch đơn, RNA mạch kép, DNA mạch kép và DNA mạch đơn.

Vi khuẩn thuộc nhóm sinh vật *prokaryote*, đơn bào chưa có nhân thật (chưa có màng nhân phân biệt nhân với tế bào chất). Bộ gen của vi khuẩn cấu tạo từ DNA mạch kép, gồm gen nhân và gen trong các plasmid. Gen nhân cấu tạo chủ yếu từ một hoặc một số phân tử DNA mạch kép dạng sợi, dạng vòng hoặc cả dạng sợi và dạng vòng. Gen nhân có thể là 1 phân tử DNA dạng vòng hoặc dạng sợi gọi là một nhiễm sắc thể đơn (vi khuẩn *E. Coli*...), hoặc gồm 2 - 3 phân tử DNA dạng vòng, dạng sợi hoặc vừa dạng vòng vừa dạng sợi (ví dụ, ở vi khuẩn *Agrobacterium tumefaciens*, vi khuẩn *Burkholderia pseudomallei*...). Gen ngoài nhân của vi khuẩn gồm một số ít gen nằm trong plasmid.

Sinh vật bậc cao có bộ gen gồm gen nhân và gen ngoài nhân. Gen nhân cấu tạo từ các phân tử DNA mảnh kép, liên kết với các phân tử protein kiềm (histon) tập hợp thành các cấu trúc đặc thù, bắt màu các thuốc nhuộm đặc hiệu gọi là các nhiễm sắc thể. Số lượng, kích thước và hình dạng nhiễm sắc thể đặc trưng riêng cho từng loài. Gen ngoài nhân gồm một số gen nằm trong các ti thể của tế bào động vật hoặc trong các lạp thể ở tế bào thực vật.

## II. Lược sử nghiên cứu bộ gen và giải trình tự bộ gen

Tuỳ theo mục đích và phương pháp nghiên cứu bộ gen, người ta đã xây dựng được các loại bản đồ di truyền bộ gen khác nhau: bản đồ hình thái nhiễm sắc thể, bản đồ di truyền tế bào, bản đồ di truyền liên kết, bản đồ lai phóng xạ, bản đồ trình tự gen... Trong đó, bản đồ trình tự gen có độ chính xác cao nhất, được ứng dụng rộng rãi trong thực tiễn phục vụ lợi ích con người. Giải trình tự bộ gen của các loài sinh vật và lập bản đồ trình tự gen, không chỉ có ý nghĩa trong nghiên cứu nguồn gốc tiến hoá của các loài sinh vật, còn có vai trò lớn trong phát hiện các gen hỏng, gen bệnh của con người và sinh vật, là cơ sở khoa học để xuất các phương pháp điều trị bệnh có hiệu quả cho con người, vật nuôi và cây trồng.

Năm 1977 lần đầu tiên bộ gen của thực khuẩn thể φ X 174 được xác định trình tự. Đến nay, đã có hàng trăm bộ gen virus, vi khuẩn và bộ gen một số loài sinh vật bậc cao được xác định trình tự.

Năm 1975, F. Sanger và A. Maxam and W. Gilbert bằng các nghiên cứu độc lập đã phát minh phương pháp giải trình tự gen. Năm 1977 bộ gen thực khuẩn thể (Phage φX 174) được giải trình tự hoàn chỉnh (F. Sanger và cộng sự, 1977).

Năm 1981, bộ gen ti thể người được giải trình tự hoàn chỉnh.

Năm 1984, bộ gen virus *Epstein-Barr* được giải trình tự.

Năm 1995, bộ gen vi khuẩn *Mycoplasma genitalium* và vi khuẩn *Haemophilus influenzae* được giải trình tự (Venter, Fraser và Smith).

Năm 1996, xác định trình tự bộ gen nấm men (*Saccharomyces cerevisiae*).

Năm 1997, trình tự bộ gen vi khuẩn *E. coli* được xác định (Blattner,

Plunkett và cộng sự).

Năm 1998, xác định trình tự bộ gen giun tròn (*Caenorhabditis elegans*), (Sulston, Waterston và cộng sự).

Năm 1999, bộ gen ruồi dám (*Drosophila melanogaster*) được giải trình tự (Công ty Celera và cộng sự.)

Năm 2000, trình tự bộ gen cây mù tạt Mustard (*Arabidopsis thaliana*) được xác định (Dự án bộ gen người - HGP và công ty Celera).

Năm 2002, bộ gen muỗi *Anopheles gambiae* được giải trình tự.

Năm 2003, bộ gen người đã được giải trình tự hoàn chỉnh.

Năm 2005, hoàn chỉnh giải mã bộ gen cây lúa, đã phát hiện ở lúa nước có tổng số 37 544 gen.

### **III. Bộ gen một số nhóm sinh vật**

#### *I. Kích thước bộ gen và số lượng gen của một số đại diện điển hình*

Bảng 1. 1. Kích thước bộ gen của virus, prokaryote và eukaryote

Đối tượng	Nhóm	Kích thước	Số gen
<b>Virus</b>			
Epstein-Barr (EBV)	Virus	172 282 bp	80
HIV	Virus	9749 ribonu	9
HCMV	Virus	229 000 bp (S)	200
Phage φX 174	Bacteriophage	5 386 nu (V)	10
Phage T4	Bacteriophage	172 000 bp S/V)	300
<b>Prokaryote</b>			
<i>H. influenzae</i>	Vi khuẩn	1 830 137 bp (V)	1 743
<i>M. pneumoniae</i>	Vi khuẩn	816 394 bp	680
<i>Helicobacter pylori</i>	Vi khuẩn	1 667 867 bp	1 589
<i>Vibrio cholerae</i>	Vi khuẩn	4 033 460 bp	3 890

<i>Escherichia coli</i>	Virus khuẩn	4 639 221 bp (V)	4 406
<i>Bacillus subtilis</i>	Virus khuẩn	4 214 814 bp	4 799
<i>M. tuberculosis</i>	Virus khuẩn	4 115 291 bp	3 924
<i>Burkholderia pseudomallei</i>	Virus khuẩn	3 600 000 bp (S) 2 400 000 bp (S)	
<i>A. tumefaciens</i>	Virus khuẩn	2 100 000 bp (S) 3 000 000 bp (V)	

**Eukaryote**

<i>S. cerevisiae</i>	Nấm men	$12,1 \times 10^6$ bp (S)	
<i>Streptomyces</i>	Nấm sợi	$10,0 \times 10^6$ bp (S)	?
<i>C. elegans</i>	Giun tròn	$95,5 \times 10^6$ bp (S)	19 099
<i>Arabidopsis thaliana</i>	Thực vật	$1,17 \times 10^9$ bp (S)	25 000
<i>D. melanogaster</i>	Ruồi giấm	$1,8 \times 10^8$ bp (S)	13 601
<i>Lúa mì Triticum</i>	Thực vật	$16,0 \times 10^9$ bp (S)	30 000
<i>Homo sapiens</i>	Người	$3,2 \times 10^9$ bp (S)	34 000
Ti thể <i>S. cerevisiae</i>	Nấm men	$78 \times 10^3$ bp (V)	34
Ti thể người	Người	16 569 bp (V)	37

(V = dạng vòng; S = dạng sợi )

**2. Cấu trúc bộ gen****2.1. Đặc điểm di truyền và cấu trúc bộ gen virus****2.2.1. Đặc điểm chung**

Virus là nhóm thể sống chưa có cấu tạo tế bào, bộ gen virus cấu tạo từ các phân tử DNA hoặc RNA trần, ở dạng vòng hoặc dạng sợi. Bộ gen virus gồm các dạng cơ bản sau:

Virus có bộ gen cấu trúc từ phân tử DNA mảnh kép: Adenovirus, Herpesvirus, Hepatitis B virus (VR viêm gan B - HBV), phage λ, phage T kí sinh vi khuẩn...

Virus có bộ gen cấu tạo từ DNA mảnh đơn, thường gặp ở các virus kí sinh vi khuẩn như phage φ X174, phage M 13...

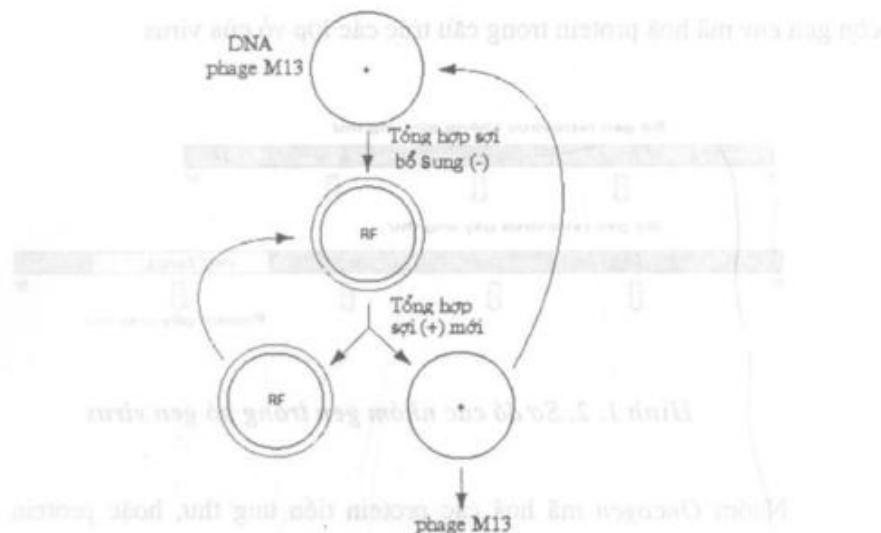
Virus có bộ gen cấu tạo từ RNA mảnh đơn, là nhóm phổ biến nhất. Nhiều loài virus có bộ gen RNA mảnh đơn như virus sởi (paramyxovirus), virus đại (rhabdovirus), virus cúm, quai bị (myxovirus influenzae) virus sốt xuất huyết, viêm não (arbovirus), virus HIV và các virus gây bệnh viêm gan: HAV, HCV, HDV, HEV...

Bộ gen virus cấu tạo từ RNA mảnh kép rất ít gặp, chủ yếu là nhóm Reovirus, gây bệnh sốt luội xanh ở cừu và một số loài động vật móng guốc.

Bộ gen virus DNA (RNA) có thể ở dạng mảnh thẳng hoặc mảnh vòng. Bộ gen virus có kích thước nhỏ khoảng 7 - 40 kb. Số lượng gen trong bộ gen rất ít, trung bình khoảng một vài chục gen, một số rất ít virus có bộ gen gồm vài trăm gen. Trong bộ gen virus có những vùng gen có khả năng biến đổi mạnh (gọi là vùng siêu biến đổi - hypervariable region), tính chất này giúp virus thích ứng nhanh với điều kiện môi trường thay đổi.

Đặc điểm tái bản bộ gen virus mang tính đặc trưng riêng ở mỗi loài, tuỳ theo vật chất di truyền của bộ gen là DNA hay RNA. Các virus có bộ gen cấu tạo từ DNA mảnh kép như các phage T, phage λ, adenovirus, virus viêm gan B (HBV)... bộ gen tái bản theo cơ chế bán bảo thủ, kiểu Okazaki.

Virus có bộ gen DNA mảnh đơn như phage φ X174, phage M 13... , bộ gen tái bản từ các dạng tái bản RF (Replicating Form). Sau khi xâm nhiễm đưa DNA vào tế bào chủ, sợi đơn DNA của phage nối thành vòng gọi là sợi dương (+). Sợi dương (+) được dùng làm khuôn để tổng hợp sợi âm bổ sung (-) theo nguyên lý Chargaff tạo thành vòng kép gọi là dạng tái bản (RF). Dạng tái bản RF có thể tái bản tạo nên hàng loạt RF mới, hoặc sử dụng sợi âm (-) của RF làm khuôn để tổng hợp nên các sợi đơn DNA (sợi +) của bộ gen phage.

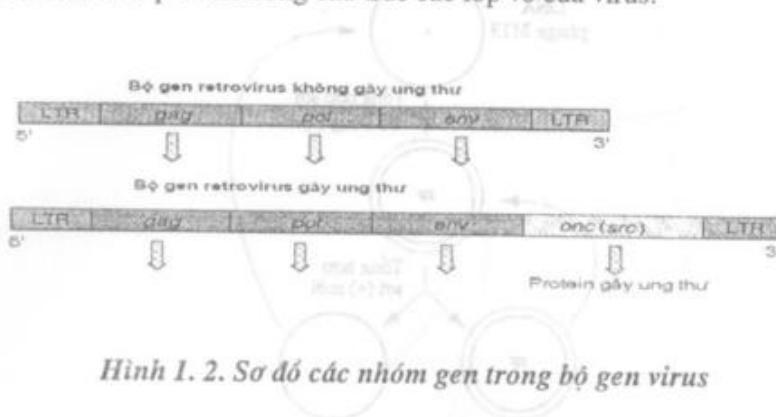


**Hình 1.1. Cơ chế tái bản DNA của phage M13**

Virus RNA tái bản bộ gen nhờ hoạt động của enzym phiên mã ngược của virus (reverse transcriptase). Sau khi RNA của virus được đưa vào tế bào, có quá trình phiên mã và dịch mã sớm một số gen mã hoá enzym phiên mã ngược và các enzym cần thiết khác của virus. Enzym phiên mã ngược của virus xúc tác tổng hợp một mạch đơn DNA bổ sung (cDNA - complement DNA) tạo phân tử lai RNA-cDNA, mạch đơn cDNA được sử dụng làm khuôn để tái bản bộ gen RNA của virus. Một số virus sử dụng enzym phiên mã ngược, enzym DNApolymerase tổng hợp cDNA mạch kép, cDNA mạch kép gắn vào bộ gen tế bào vật chủ tạo nên các provirus (tiền virus) tồn tại cùng tế bào chủ. Trong những điều kiện nhất định (nhiệt độ cao, chất độc...) provirus tách khỏi bộ gen tế bào chủ, một trong 2 mạch cDNA được sử dụng làm khuôn để tái bản bộ gen RNA của virus.

Bộ gen virus có kích thước nhỏ, số lượng gen ít, cấu trúc đơn giản. Ví dụ, retrovirus có bộ gen RNA sợi đơn, gồm hai sợi RNA cấu trúc và kích thước giống nhau, kích thước khoảng 7 - 11 kb, gồm nhóm gen chủ yếu: *gag*, *pol*, *env*. Một số virus gây ung thư có thêm nhóm gen tiền ung thư SRC- *oncogen*. Trong bộ gen virus, nhóm gen *gag* mã hoá protein lõi trong thành phần cấu trúc hạt virion, gen *pol* mã hoá enzym phiên mã ngược và các enzym cần thiết cho hoạt động của virus,

còn gen *env* mã hoá protein trong cấu trúc các lớp vỏ của virus.



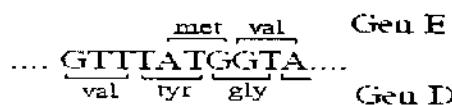
Hình 1.2. Sơ đồ các nhóm gen trong bộ gen virus

Nhóm *Oncogen* mã hoá các protein tiền ung thư, hoặc protein gây hiện tượng phân chia không kiểm soát của tế bào, tạo nên các khối u ở cơ thể sinh vật bậc cao. Hai đầu mỗi sợi RNA có một trình tự lặp dài (LTR - Long Terminal Repeats). Đoạn LTR mang một số trình tự cần thiết cho quá trình biểu hiện gen của virus và một trình tự đặc hiệu – *att* (trình tự gắn xen- attachment site), giúp cho bộ gen virus gắn với bộ gen tế bào chủ. Mặt khác, trình tự LTR có khả năng hạn chế tác động của các enzym nuclease trong tế bào.

#### 2.2.2. Đặc điểm cấu trúc bộ gen một số virus điển hình

##### a. Cấu trúc bộ gen phage φ X 174

Thực khuẩn thể φ X174 (phage φ X174) xâm nhiễm tế bào vi khuẩn *E. coli*, có bộ gen DNA một mạch đơn, dạng vòng. Bộ gen φ X174 được giải mã hoàn chỉnh năm 1977 (F. Sanger và cộng sự). Bộ gen phage φ X174 kích thước nhỏ, gồm 5 386 nucleotid, có khoảng 10 gen. Bộ gen gồm phần lớn trình tự DNA mang mã di truyền. Khung đọc mã của các gen khác nhau tạo nên hiện tượng mã trùm lên nhau (overlapping gen) và gen này nằm trong gen khác. Ví dụ, trong bộ gen φ X174 có gen E với bộ 3 mã mở đầu giữa một bộ 3 của gen D, hoặc gen B nằm trong gen A...



Giữa gen A và gen H là đoạn trình tự DNA gồm 66 nucleotid. Giữa gen F và gen J là gen đệm dài 39 nucleotid, giữa gen F và gen G là đoạn đệm (spacer) dài 111 nucleotid. Các gen A, B, C và D chỉ huy quá trình tái bản DNA của phage. Gen E mã hoá tổng hợp enzym lysozym làm tan tế bào vi khuẩn. Các gen J, F, G và gen H mã hoá các protein cấu tạo vỏ phage. Hiện tượng gen nằm trong gen và mã trùng nhau cũng gặp trong bộ gen của nhiều sinh vật bậc cao và hệ gen tì thể người.

### b. Cấu trúc bộ gen HIV

HIV là virus gây hội chứng suy giảm miễn dịch ở người (HIV - Human Immuno - deficiency Virus). Bộ gen HIV gồm hai phân tử RNA kích thước 9749 ribonucleotid, gồm 9 gen chính *gag*, *pol*, *env*, *vif*, *tat*, *vpu*, *nef*, *vpr*, *rev*. Các gen có kích thước và chức năng khác nhau, có hiện tượng các gen trùng lén nhau, gen này nằm trong cấu trúc của gen khác và hiện tượng gen không liên tục (một gen gồm nhiều trình tự ở các vị trí khác nhau trên RNA như gen *tat*, *rev*...) hoặc khung đọc mã (ORF) khác nhau nên cùng một trình tự tạo nên các gen khác nhau.

Chức năng các gen của HIV gồm:

Gen *gag* mã hoá các protein capsid (p17 - protein capsid phụ; p24 - protein capsid chính; p7, p9 - protein nucleocapsid bám dính).

Gen *pol* mã hoá các protein enzym (p10 - enzym protease; p66/51 - enzym phiên mã ngược (reverse transcriptase); p32 - enzym gắn xen (integrase))

Gen *env* mã hoá các protein vỏ (gp120 - glycoprotein vỏ ngoài; gp41 - glycoprotein lớp màng trong)

Gen *tat*: Điều chỉnh sự phiến mã tích cực của virus

Gen *rev*: Điều hòa sự biểu hiện gen virus

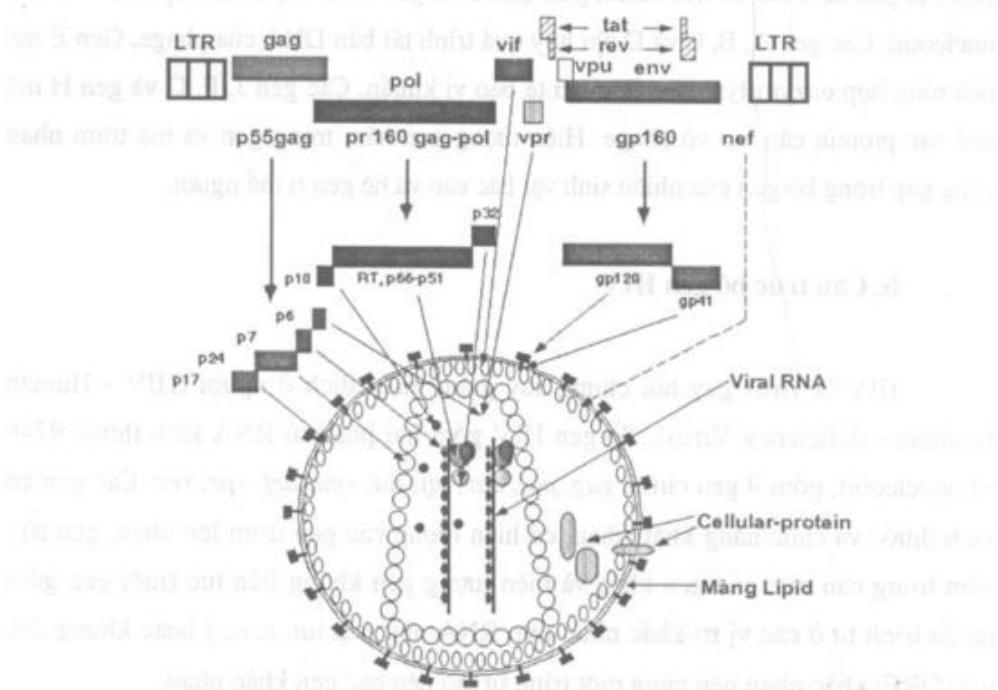
Gen *vif*: Protein quy định tính lây nhiễm của virus

Gen *vpu*: Viral protein U, tham gia giải phóng virus khỏi tế bào

Gen vpr: Viral protein R, bắt hoạt gen tế bào chủ, làm suy yếu tế bào chủ

Gen nef: Vô hiệu hoá sự điều chỉnh phiền mā

LTR (Long Terminal Repeats) - trình tự lặp dài

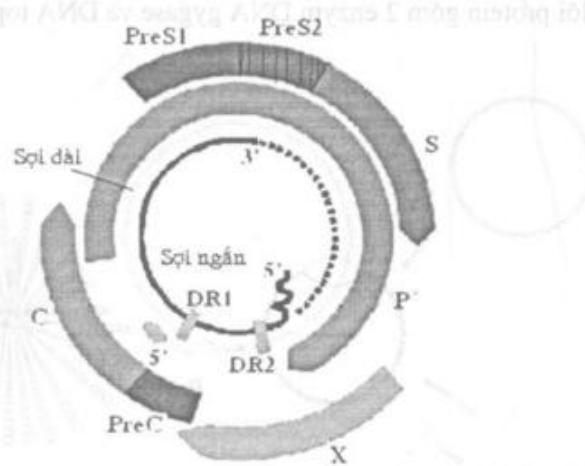


Hình 1.3. Bộ gen HIV -I và cấu trúc virion

### c. Cấu trúc bộ gen virus viêm gan B (HBV)

Bộ gen HBV là phân tử DNA xoắn kép không hoàn toàn, gồm 2 mạch DNA có độ dài khác nhau. Sợi đơn phía ngoài là khuôn tổng hợp các mRNA có kích thước 3200 nucleotid gọi là sợi dài (L), sợi bên trong nhỏ hơn có kích thước khoảng 2400 nucleotid gọi là sợi ngắn (S). Bộ gen HBV bị đột biến tạo nên nhiều biến thể khác nhau. Bộ gen HBV mã hóa 4 nhóm protein: Nhóm gen mã hóa kháng nguyên HbsAg. Kháng nguyên bê mặt HbsAg gồm 3 loại khác nhau do 3 phân đoạn gen S mã hóa: phân đoạn PreS1 mã hóa chuỗi peptid gồm 108 aa, phân đoạn

PreS2 mã hoá chuỗi peptid 55 aa và gen S mã hoá chuỗi peptid gồm 226 aa. Gen C mã hoá kháng nguyên HbcAg và protein C (gồm PreC và gen C - 183 aa). Gen P là gen có kích thước lớn nhất, mã hoá protein gồm 832 aa, gồm các enzym DNA polymerase cần thiết cho quá trình tái bản DNA của virus HBV. Gen X có kích thước nhỏ nhất, mã hoá HBx protein có 154 aa, là các protein điều hoà quá trình dịch mã của virus HBV.



Hình 1. 4. Cấu trúc bộ gen virus HBV

## 2. 2. Đặc điểm di truyền và cấu trúc bộ gen prokaryote

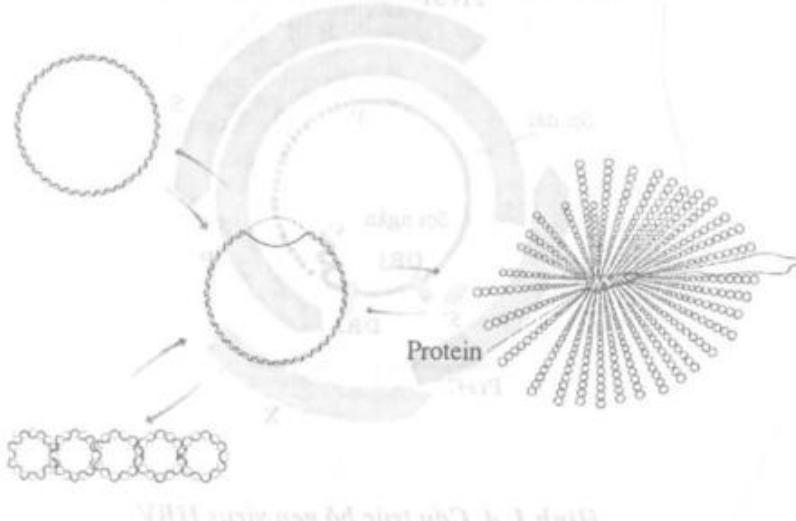
### 2. 2. 1. Đặc điểm chung

Sinh vật *prokaryote* điển hình là các loài vi khuẩn, có bộ gen cấu tạo từ DNA xoắn kép dạng vòng, hoặc dạng sợi. Một số loài vi khuẩn có bộ gen gồm hai hay nhiều phân tử DNA đều dạng sợi, dạng vòng hoặc có dạng vòng và dạng sợi. Ví dụ, vi khuẩn *Agrobacterium tumefaciens* có bộ gen gồm một phân tử DNA dạng sợi kích thước 2 100 000 bp, một phân tử DNA dạng vòng kích thước 3 000 000 bp; vi khuẩn *Rhodobacter sphaeroides* có bộ gen gồm 2 phân tử DNA dạng vòng, kích thước 3,0 Mb và 0,7 Mb, vi khuẩn *Paracoccus denitrificans* có bộ gen gồm 3 phân

từ DNA dạng vòng, kích thước 2,0 Mb ; 1,1 Mb và 0,64Mb...

Bộ gen *prokaryote* tương đối đồng nhất, phần lớn các gen trong bộ gen mang mã di truyền. Phần lớn các gen trong bộ gen *E. coli* mang mã di truyền, chỉ khoảng 11% không mang mã (Brown, 2000).

Bộ gen vi khuẩn có thể chuyển hóa giữa các dạng vòng đơn, vòng xoắn hoặc siêu xoắn. Ví dụ, vi khuẩn *E. coli* có bộ gen gồm 40 - 50 vòng siêu xoắn quanh phần lõi protein gồm 2 enzym DNA gyrase và DNA topoisomerase.



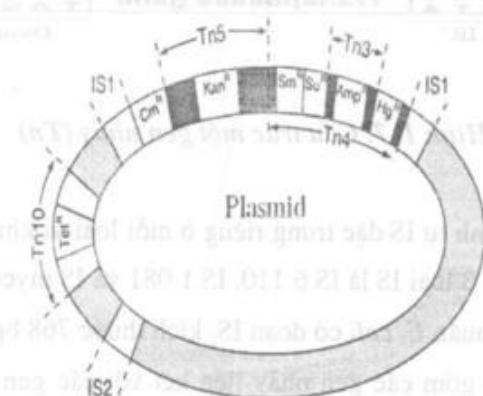
Hình 1.5. Sơ đồ các dạng cấu trúc bộ gen *E. coli*

Trong bộ gen, các gen cùng hướng trao đổi chất hợp thành các Operon, giúp vi khuẩn thích ứng nhanh với điều kiện môi trường thay đổi. Operon là cấu trúc đặc trưng của bộ gen của *prokaryote*. Operon gồm một số gen cùng hướng trao đổi chất, tập hợp thành cụm gần nhau, cùng chung một vùng 5' UTR và vùng 3'UTR. Ví dụ, vi khuẩn *E. coli* có operon Lac gồm 3 gen *Lac Z*, *Lac Y* và *Lac A*; Operon Tryptophan gồm 5 gen: *Trp E*, *Trp D*, *Trp C*, *Trp B* và *Trp A*. Cấu trúc operon giúp vi khuẩn thích ứng nhanh với điều kiện môi trường thay đổi. Số lượng operon trong bộ gen của các loài vi khuẩn mang tính đặc trưng riêng. Ví dụ, vi khuẩn *E. coli* bộ gen có kích thước 4 639 221 bp, gồm 4 397 gen, gồm hơn 600

operon, bộ gen vi khuẩn *Bacillus subtilis* có hàng trăm operon...

**Đoạn 01:** Bộ gen vi khuẩn có rất ít trình tự DNA lặp lại, phần lớn DNA trong bộ gen mang mã di truyền mã hoá các protein (tới 90%, ở người < 3%). Trong bộ gen vi khuẩn cũng có một số trình tự của bộ gen bacteriophage. Ví dụ, trong bộ gen *E. coli* có 8 chỗ trong bộ gen mang trình tự đặc trưng của bacteriophage.

**Đoạn 02:** Các gen trong plasmid là thành phần quan trọng của bộ gen vi khuẩn. Mỗi tế bào vi khuẩn trong điều kiện bình thường có khoảng 15 - 20 plasmid dạng vòng, kích thước plasmid có thể từ 1- 13 kb. Ví dụ, vi khuẩn *Borrelia burgdorferi* gồm gen nhân là phân tử DNA dạng sợi, kích thước 910 kb (mang 853 gen) và 17 plasmid có tổng kích thước tới 533 kb (có 430 gen phân bố trên 11 plasmid). Trong tế bào vi khuẩn có nhiều loại plasmid khác nhau: plasmid giới tính, plasmid kháng chất kháng sinh, plasmid độc tố..... Trong môi trường có chất kháng sinh, độc tố... số lượng plasmid trong tế bào vi khuẩn có thể tăng lên tới hàng nghìn plasmid trong một tế bào. Đặc tính này giúp vi khuẩn chống lại các yếu tố bất lợi của môi trường, gây các hiện tượng nhòn thuốc kháng sinh hoặc kháng chất kháng sinh...

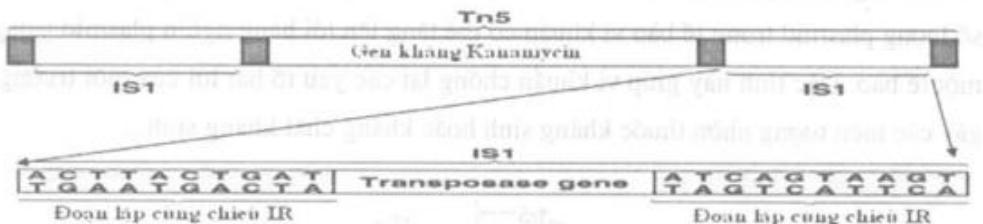


Hình 1.6. Plasmid và các gen kháng chất kháng sinh ở *E. coli*

Trong bộ gen của prokaryote còn có các gen nhảy (Transposon - Tn). Gen nhảy có thể nằm trên DNA bộ gen hoặc trên DNA của các plasmid. Gen nhảy có kích thước, tính chất đặc trưng riêng ở mỗi loài vi khuẩn.

Ví dụ, vi khuẩn *E. coli* có gen nhảy Tn9 dài 1 100 bp, mang gen kháng chloramphenicol; Gen nhảy Tn5 mang gen kháng kanamycin; Gen nhảy Tn10 kích thước 9 kb, mang gen mã hoá *Tetracyclin*; Gen nhảy Tn 1 681 mang gen sinh độc tố gây bệnh tiêu chảy... các gen nhảy thường nằm trên các plasmid. Plasmid có thể có thể có một hoặc một số gen nhảy.

Gen nhảy được phân biệt hai dạng. Dạng thứ nhất gồm trình tự di truyền vận động TG (transposase gene) nằm giữa hai đoạn lặp cùng chiều IR (Inverted Repeat) kích thước khoảng 8 - 40 bp, gọi là đoạn xen IS (insert sequence). Đoạn xen IS kích thước khoảng 800 - 1500 bp, có thể thay đổi vị trí trên bộ gen vi khuẩn. Khi gắn xen vào bộ gen hoặc plasmid, gây biến đổi trạng thái của bộ gen. Ngược lại, khi tách khỏi bộ gen, bộ gen trở lại trạng thái bình thường.



Hình 1.7. Cấu trúc một gen nhảy (Tn)

Kích thước trình tự IS đặc trưng riêng ở mỗi loài vi khuẩn. Ví dụ, vi khuẩn lao *M. tuberculosis* có 3 loại IS là IS 6 110, IS 1 081 và IS myco kích thước khoảng 968 đến 1381 bp; vi khuẩn *E. coli* có đoạn IS<sub>1</sub> kích thước 768 bp.

Dạng thứ hai, gồm các gen nhảy liên kết với các gen mã hoá chất kháng sinh, gen sinh độc tố, hoặc gen kháng chất kháng sinh... Số lượng gen nhảy, kích thước của mỗi loại gen nhảy đặc trưng riêng ở mỗi loài vi khuẩn. Có thể sử dụng tính đặc trưng về kích thước và trình tự của đoạn IS, làm cơ sở chẩn đoán nhanh các bệnh nhiễm khuẩn bằng kỹ thuật PCR.

### 2.2.2. Đặc điểm cấu trúc bộ gen một số đại diện prokaryote

#### a. Cấu trúc bộ gen vi khuẩn *Escherichia coli* K12

Bộ gen vi khuẩn *Escherichia coli* K12 là một phân tử DNA xoắn kép dài khoảng 1,6 mm tạo thành 40 - 50 vòng siêu xoắn, mỗi vòng kích thước khoảng 100 kb. Bộ gen vi khuẩn *Escherichia coli* K12 được giải mã hoàn thiện năm 1977. Trong bộ gen *E. coli*, gần 89% hàm lượng DNA mang mã di truyền, gồm 4285 gen mã hoá protein, 122 gen mã hoá các dạng RNA, và các gen mã hoá các trình tự điều hoà hoạt động gen, các gen nhảy... Các gen mã hoá protein trong bộ gen của vi khuẩn *E. coli* trung bình 317 acid amin/ một chuỗi polypeptid.

Phần DNA không mang mã di truyền trong bộ gen *E. coli* chiếm khoảng 11%, bao gồm các đoạn DNA intergenic và 77 trình tự lặp lại. Các đoạn intergenic có kích thước trung bình 130 bp, đoạn có kích thước lớn nhất là đoạn lặp các trình tự DNA không mang mã, khoảng 1730 bp. Bộ gen *E. coli* K12 có khoảng 630 - 700 operon, có 1632 gen chưa rõ chức năng chiếm 38,6 % bộ gen (Blattner và cộng sự 1997).

#### b. Cấu trúc bộ gen vi khuẩn *Methanococcus jannaschii*

Bộ gen vi khuẩn *Methanococcus jannaschii* được Viện nghiên cứu Bộ gen (TIGR - The Institute for Genomic Research ) giải trình tự 1996, là vi khuẩn tự dưỡng đầu tiên được giải trình tự. Bộ gen của vi khuẩn *M. jannaschii* gồm một phân tử DNA xoắn kép dạng vòng, kích thước lớn (1 664 976 bp) và 2 phân tử DNA dạng sợi có kích thước 58 407 bp và 16 550 bp.

Bộ gen gồm 1 743 gen mang mã di truyền trong đó 1 682 gen nằm trên DNA dạng vòng. Phân tử DNA dạng sợi kích thước lớn có 44 gen, phân tử DNA kích thước nhỏ mang 12 gen. Các gen mã hóa protein thường tập hợp trong các operon.

### c. Cấu trúc bộ gen *Mycoplasma genitalium*

Bộ gen *Mycoplasma genitalium* được các nhà khoa học thuộc Trường Đại học Tổng hợp John Hopkin giải trình tự 1995. Vì khuôn *M. genitalium* là đại diện *prokaryote* có bộ gen kích thước nhỏ nhất 580 070 bp. Bộ gen gồm 486 gen mã hoá protein, chiếm 85% hàm lượng DNA bộ gen. Kích thước trung bình các gen mã hoá protein là 1040 bp. Trong bộ gen có hiện tượng mã trùng lén nhau (overlapping gene) như ở virus. Bộ gen của *M. genitalium* gồm các gen có cấu trúc đồng nhất, gồm các trình tự DNA mã hoá protein.

### 2.3. Đặc điểm di truyền và cấu trúc bộ gen eukaryote

#### 2.3.1. *Đặc điểm chung*

Sinh vật *eukaryote* có cấu tạo tế bào điển hình, nhân phân biệt với tế bào chất bởi màng nhân. Bộ gen của sinh vật *eukaryote* gồm hệ gen nhân và hệ gen ngoài nhân. Gen nhân cấu tạo từ các phân tử DNA xoắn kép, có cấu trúc phức tạp, gồm DNA kết hợp với các loại protein kiềm, hình thành nên các cấu trúc đặc trưng gọi là nhiễm sắc thể (chromosome). Nhiễm sắc thể có cấu trúc, số lượng đặc trưng riêng cho mỗi loài sinh vật. Ví dụ, ở nấm men  $2n = 16$ , ở người  $2n = 46$ , bò  $2n = 60$ , gà  $2n = 78$ , ruồi giấm  $2n = 8$ , lúa  $2n = 24$ , ngô  $2n = 20...$

Hình dạng, kích thước và số lượng nhiễm sắc thể không liên quan đến mức độ tiến hoá của các loài sinh vật. Các gen nhân phân bố trên các nhiễm sắc thể, sự di truyền các gen nhân có tính qui luật, tương đối chặt chẽ đảm bảo thông tin di truyền được di truyền và ổn định tương đối từ thế hệ này sang thế hệ khác.

Hệ gen ngoài nhân của sinh vật *eukaryote* chủ yếu tập trung trong các ti thể đối với tế bào động vật và nấm men, hoặc trong lạp thể đối với tế bào thực vật. Hệ gen ngoài nhân gồm các phân tử DNA trần, dạng vòng hoặc dạng sợi. Ở người trong mỗi tế bào có khoảng 8 000 ti thể, trong mỗi ti thể có mười phân tử DNA dạng vòng giống nhau, mỗi phân tử DNA có kích thước 16 569 bp, gồm 37 gen

trong đó có 13 gen mã hoá protein, 22 gen mã hoá tRNA và 2 gen mã hoá rRNA. Tế bào nấm men có khoảng 6 500 ti thể, mỗi ti thể chứa một phân tử DNA xoắn kép dạng vòng, kích thước 75 000 bp gồm 35 gen. Hàm lượng DNA ti thể của tế bào nấm men lớn gấp 100 lần hàm lượng DNA trong nhân tế bào nấm men.

Trong mỗi tế bào thực vật có nhiều lạp thể, nên hàm lượng DNA trong các lạp thể rất lớn. Ví dụ, ở tảo lục (*Chlamydomonas*) mỗi tế bào có khoảng 1 000 lạp thể, mỗi lạp thể có hệ gen kích thước khoảng 195 000 bp; mỗi lạp thể trong tế bào lá cây thuộc lá có hệ gen lạp thể với kích thước 16 500 bp.

Bộ gen của sinh vật *eukaryote* gồm DNA mang mã di truyền mã hoá protein hoặc RNA chiếm một phần nhỏ trong bộ gen, và các DNA không có chức năng hoặc chưa rõ chức năng (intron và các đoạn dị nhiễm sắc) chiếm hàm lượng lớn trong bộ gen. Ví dụ, ở người có dưới 2% DNA trong bộ gen mã hóa các protein, phần còn lại tập trung trong các đoạn lặp DNA, đoạn dị nhiễm sắc...

Bộ gen của sinh vật *eukaryote* có thể chia làm ba nhóm DNA: nhóm DNA lặp lại nhiều, nhóm DNA lặp lại ít và nhóm DNA không lặp lại.

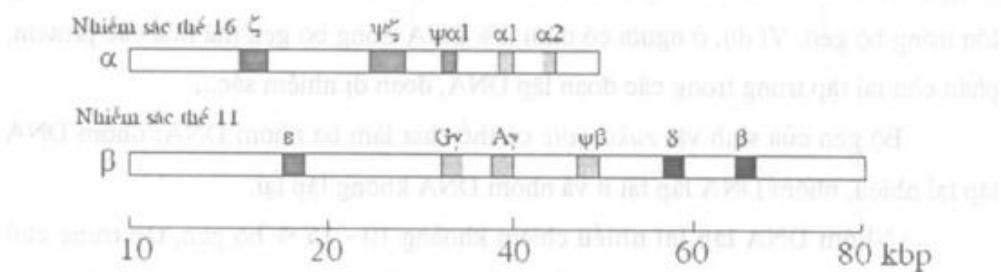
**Nhóm DNA lặp lại nhiều** chiếm khoảng 10 - 15 % bộ gen, tập trung chủ yếu quanh tâm động (trình tự CEN), hoặc các đầu mút nhiễm sắc thể (trình tự TEL). DNA lặp lại nhiều tạo thành những đoạn dị nhiễm sắc (heterochromatine), VNTR, SSR...

Các đoạn DNA lặp lại VNTR (Variable Number of Tandem Repeat) gồm DNA minisatellit và DNA microsatellit. DNA minisatellit là những đoạn DNA lặp lại, có kích thước 1-5 kb, lặp lại nhiều lần một trình tự 14 - 500 bp phân bố khắp bộ gen. DNA microsatellit là các đoạn DNA lặp kích thước khoảng 100 kb, lặp lại nhiều lần một trình tự ngắn 2-6 bp. Trong mỗi bộ gen người thường có khoảng  $10^6$  bản sao microsatellit, tập trung chủ yếu quanh tâm động của các nhiễm sắc thể, tạo nên các vùng dị nhiễm sắc (heterochromatin).

Trình tự telomer còn gọi là trình tự TEL là những đoạn lặp ở cuối nhiễm sắc thể, thường là các trình tự ngắn 6 - 9 bp lặp lại từ 250 – 1 000 lần. Ví dụ, trong tế bào người trình tự TEL là đoạn lặp nhiều lần một trình tự 6 nucleotid TAAGGG;

ở cây mutard (*Arabidopsis*) trình tự TEL là sự lặp nhiều lần một trình tự 6 TTTAGGG... (nhóm DNA lặp lại) và ở con người là trình tự 5' GGGGCCCC 3'.

Nhóm DNA lặp lại ít gồm DNA trong các họ gen (họ gen globin, actin, họ gen mã hoá rRNA, tRNA...). DNA trong họ gen mã hoá tRNA (lặp lại từ 10 - 50 lần trong mỗi tế bào), các gen mã hoá rRNA (lặp lại từ 200 - 250 lần trong mỗi tế bào)... Bộ gen người có các họ gen mã hoá protein và các họ gen mã hoá rRNA, họ gen mã hoá protein điển hình là họ gen globin, họ gen myosin và họ gen actin... Trong bộ gen người, họ gen globin tập hợp chủ yếu trên nhiễm sắc thể số 16 và nhiễm sắc thể số 11. Trên nhiễm sắc thể 16 tập hợp các gen mã hoá  $\alpha$ -globin, trên nhiễm sắc thể số 11 tập hợp các gen mã hoá  $\beta$ -globin.



Hình 1.8. Họ gen globin trong bộ gen người

Họ gen mã hoá rRNA trong bộ gen người nằm chủ yếu ở các nhiễm sắc thể số 1, 13, 14, 15, 21 và 22. Các gen mã hoá rRNA 28S, 18S và 5, 8S tập hợp thành 5 nhóm, mỗi nhóm được lặp lại 50 - 70 lần trên các nhiễm sắc thể số 13, 14, 15, 21 và 22. Trên nhiễm sắc thể số 1 có khoảng 2 000 gen mã hoá 5S rRNA.

Các trình tự SINE (Short interspersed elements) là các trình tự Alu, hoặc các đoạn gồm từ 200 - 300 bp lặp lại từ 100 000 đến 300 000 lần...

Các trình tự LINE (Long interspersed elements) là các trình tự từ 1 - 5 kb, lặp lại 10 lần đến 100 000 lần trong bộ gen.

Các gen giả (pseudogenes) là những trình tự DNA có cấu trúc tương tự như các gen mã hoá protein nhưng không được phiên mã, hoặc được phiên mã nhưng

không được dịch mã.

**Nhóm DNA không lặp lại** là thành phần DNA trong các gen mã hoá protein, enzym. Nhóm DNA không lặp lại ở các sinh vật tiến hoá càng cao, chiếm tỉ lệ càng thấp trong bộ gen.

Bộ gen của sinh vật *eukaryote* gồm các gen có chức năng riêng, không có hiện tượng tạo thành các operon. Các gen mã hoá protein ở *eukaryote* có sự xen kẽ giữa exon và intron (gen phân đoạn), trừ các gen mã hoá histon không có intron. Số lượng intron và kích thước intron khác nhau ở mỗi gen và phụ thuộc tính đặc trưng của loài. Ví dụ, trong bộ gen người gen mã hoá β - globin kích thước 1 660 bp trong đó intron gồm 670 bp chiếm 40% kích thước gen; Gen mã hoá factor VIII có kích thước khoảng 186 000 bp, gồm exon gồm 9 000 bp chiếm khoảng 5% và intron có kích thước khoảng 177 000 bp, chiếm gần 95 % kích thước gen.

Gen mã hoá protein của sinh vật bậc cao có cấu trúc tương đối phức tạp gồm trình tự điều hoà hoạt động gen gồm: promoter, operator, enhancer, silencer... vùng mang mã di truyền mã hoá protein, enzym và vùng kết thúc mang tín hiệu kết thúc, tín hiệu adenyl hoá và điểm mút (telomer).

### 2.3.2. Đặc điểm cấu trúc bộ gen một số đại diện *eukaryote*

#### a. Bộ gen nấm men (*Saccharomyces cerevisiae*)

Bộ gen nấm men có kích thước 13500 kp, tập hợp thành 16 nhiễm sắc thể. Các nhiễm sắc thể của nấm men có kích thước khác nhau: nhiễm sắc thể số IV có kích thước lớn nhất (1352 kb), nhiễm sắc thể số 1 có kích thước nhỏ nhất (230 kb). Bộ gen nấm men gồm khoảng 5 885 gen mã hoá protein (coding gene), gần 140 gen mã hoá RNA ribosomal (rRNA), 40 gen mã hoá snRNA (small nuclear RNA) và 275 mã hoá RNA vận chuyển (tRNA). Bộ gen nấm men có mật độ gen mã hoá protein (coding gene) cao hơn rất nhiều so với các đại diện *eukaryote* khác (*Caenorhabditis elegans*, *Drosophila melanogaster*, người...).

Trong bộ gen nấm men, số lượng gen có intron chiếm tỉ lệ rất thấp (khoảng

thường có kích thước nhỏ, các trình tự DNA lặp lại chiếm tỉ lệ ít hơn rất nhiều so với các sinh vật *eukaryote* khác. Trong tổng số 5 885 gen mã hoá protein của tế bào nấm men đã biết chính xác 3 408 loại protein, trong đó hơn 1 000 loại protein giống như ở sinh vật bậc cao khác. Ở nấm men, nhiều gen có cấu trúc tương đồng với các gen của sinh vật *prokaryote*.

### b. Bộ gen giun tròn (*Caenorhabditis elegans*)

Bộ gen giun tròn *C. elegans* có kích thước khoảng 97 Mb phân bố trên 4 cặp nhiễm sắc thể số I, II, III, IV và nhiễm sắc thể giới tính X. Sự khác biệt về giới tính ở giun tròn *C. elegans* xác định bởi cặp nhiễm sắc thể giới tính X: con cái XX, con đực XO (giun tròn không có nhiễm sắc thể Y). Bộ gen giun tròn được giải trình tự năm 1998.

Bảng 1. 2. Số lượng gen và kích thước bộ gen của *C. elegans*

Nhiễm sắc thể	Kích thước nhiễm sắc thể (Mb)	Số lượng gen mã hoá protein	Kích thước trung bình 1 gen (kb)	Số lượng gen mã hoá tRNA
I	7,9	2803	5,06 kb	13
II	8,5	3259	3,65	6
III	7,6	2508	5,40	7
IV	9,2	3094	5,17	5
X	10,1	2631	6,54	3

Bộ gen giun tròn *C. elegans* lớn gấp 8 lần bộ gen nấm men, nhưng số lượng gen mã hoá protein trong bộ gen chỉ lớn hơn ở bộ gen nấm men 3 lần (gồm 19 099 gen), tỉ lệ DNA lặp lại hơn ở bộ gen nấm men nhiều lần. Bộ gen giun tròn *C. elegans* có tần số gen tương đối thấp, trung bình khoảng 5 kb DNA / 1 gen. Phần lớn các gen đều có các intron, mỗi gen có trung bình 5 intron. Tỉ lệ DNA trong các

exon thấp, exon chiếm khoảng 27% hàm lượng DNA của bộ gen. Trong bộ gen giun tròn *C. elegans* các trình tự DNA lặp lại chiếm khoảng 2,6% hàm lượng DNA bộ gen. Các gen mã hoá ribosomal RNA phân bố chủ yếu ở đoạn DNA lặp lại trên nhiễm sắc thể số I.

#### c. Bộ gen cây mù tạt (*Mustard - Arabidopsis thaliana*)

Cây mù tạt *Arabidopsis thaliana* thuộc ngành thực vật có hoa, kích thước bộ gen 125 Mb, gồm 25 498 gen. Bộ gen cây mù tạt (*Arabidopsis thaliana*) được giải trình tự năm 2000 (Dự án Bộ gen người và công ty Celera, 2000). Cây mù tạt có 5 cặp nhiễm sắc thể. Bộ gen gồm các gen có kích thước nhỏ, một gen có kích thước trung bình 4,6 kb. Các gen có cấu trúc exon xen kẽ intron, kích thước trung bình một exon gồm khoảng 250 bp, kích thước trung bình một intron là 170 bp. Hàm lượng DNA ti thể của *Arabidopsis thaliana* chiếm gần 25% tổng lượng DNA trong tế bào (ở tế bào động vật hệ gen ti thể chiếm khoảng 5% bộ gen). Trong bộ gen cây mù tạt DNA lặp lại chiếm 58% hàm lượng DNA của bộ gen, gồm 24 đoạn dài trên 100 kb. Bộ gen cây mù tạt có nhiều họ gen, chiếm 17% hàm lượng DNA bộ gen (23 họ gen với 4 140 gen).

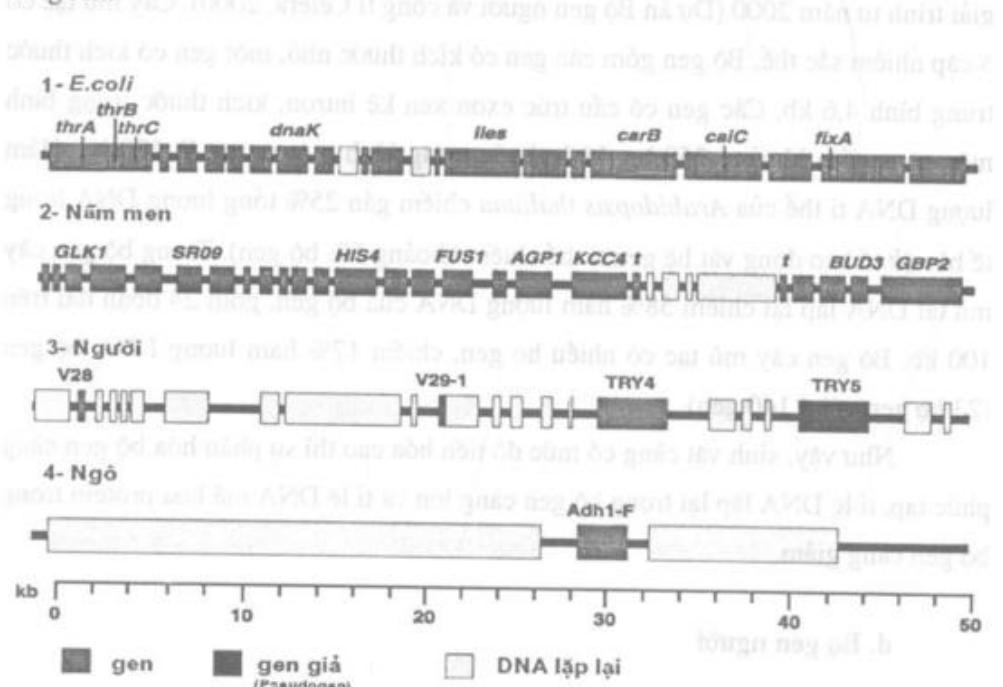
Như vậy, sinh vật càng có mức độ tiến hóa cao thì sự phân hóa bộ gen càng phức tạp, tỉ lệ DNA lặp lại trong bộ gen càng lớn và tỉ lệ DNA mã hóa protein trong bộ gen càng giảm.

#### d. Bộ gen người

Bộ gen người có cấu trúc tương đối phức tạp, gồm hệ gen nhân và gen ti thể. Gen nhân có kích thước khoảng  $3,0 \times 10^9$  nucleotid (3 000 Mb), gấp 3 lần kích thước bộ gen *C. elegans* hoặc *D. melanogaster*. Bộ gen người gồm 22 cặp nhiễm sắc thể thường (autosome) và một cặp nhiễm sắc thể giới tính, đàn ông mang cặp nhiễm sắc thể XY, đàn bà XX. Hàm lượng DNA trong các nhiễm sắc thể phân bố không đều, biến động từ 48 Mb đến 279 Mb.

Trong bộ gen người, các gen mã hoá protein có cấu trúc xen kẽ exon và intron, exon có kích thước nhỏ còn intron thường có kích thước rất lớn, mỗi gen thường có từ 1 đến hơn 75 exon. Bộ gen người có tỉ lệ gen mã hoá protein thấp (chiếm dưới 3% hàm lượng DNA bộ gen), hàm lượng DNA lặp lại chiếm tỉ lệ rất cao (trên 70% hàm lượng DNA của tế bào).

Phân tích đoạn 50 kb trong bộ gen *E. coli*, nấm men, bộ gen người và bộ gen cây ngô cho thấy hàm lượng DNA lặp lại ở *E. coli* chiếm một tỷ lệ rất nhỏ so với ở bộ gen người và bộ gen cây ngô. Các sinh vật tiến hoá càng cao, bộ gen có hàm lượng DNA càng lớn, tỉ lệ DNA lặp lại và DNA không mang mã di truyền càng lớn.

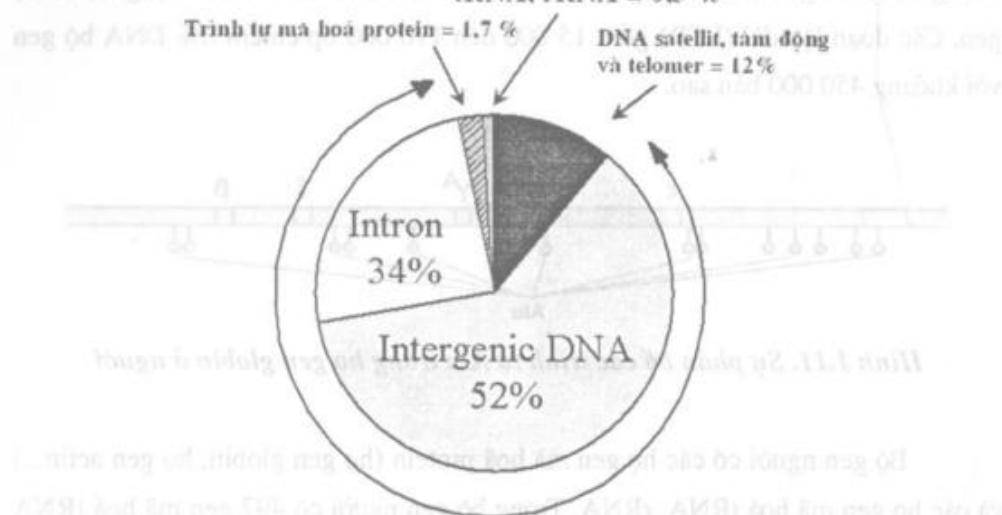


Hình 1. 9. Sự khác nhau về cấu trúc bộ gen giữa *E. coli*, nấm men,

người và cây ngô (so sánh 1 đoạn có kích thước 50 kb)

Kết quả giải mã bộ gen người cho thấy bộ gen người gồm khoảng 32000 –

40 000 gen, tỉ lệ G - C = 41%, các trình tự DNA mã hoá protein chiếm khoảng 1,7% hàm lượng DNA của bộ gen. Trong bộ gen người, hàm lượng DNA không mang mã di truyền chiếm tỉ lệ rất cao (DNA giữa các gen (intergenic) chiếm 52%, hàm lượng DNA trong các intron chiếm 34%, các trình tự DNA lặp lại không mang mã di truyền chiếm khoảng 12% hàm lượng DNA của tế bào...).



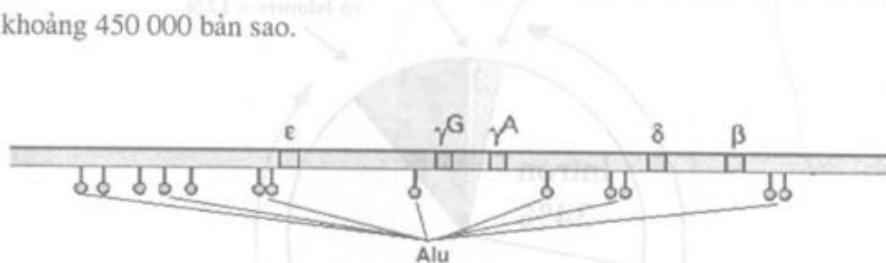
Hình 1.10. Thành phần các nhóm DNA trong bộ gen người

Bộ gen người có khoảng 40% DNA không lặp lại (single copy DNA), trong đó các gen mã hoá protein chỉ chiếm gần 3%. Các trình tự DNA lặp lại ít (từ 10 đến 105 lần) chiếm khoảng 30% bộ gen, các trình tự DNA lặp lại cao (trên 106 lần) gồm các trình tự từ 5 - 100 bp chiếm khoảng 10%. Intergenic DNA là các trình tự DNA nằm giữa các gen lặp lại nhiều lần một trình tự nhất định, bao gồm các trình tự điều hoà, promoter, enhancer..., các yếu tố di truyền vận động (transposable elements), các đoạn SINE, LINE..., và các gen giả không được dịch mã (gen Xist, gen H19, gen His-1)...

Trình tự SINE (Short Interspersed Repeats) là các đoạn DNA lặp lại ngắn,

trong tế bào người trình tự SINE điển hình là trình tự Alu. Trình tự Alu là đoạn DNA lặp nhiều lần một trình tự 100 - 300 bp. Trong bộ gen người trình tự Alu chiếm tới 13% hàm lượng DNA của tế bào, có khoảng 1500 000 bản sao các trình tự Alu phân bố khắp bộ gen, tập trung nhiều ở các họ gen. Ví dụ, họ gen globin có 13 trình tự Alu.

Trình tự LINE là các đoạn DNA lặp lại dài, kích thước 6 000 – 8 000 bp. Trong bộ gen người có khoảng 850 000 bản sao, chiếm 21% hàm lượng DNA bộ gen. Các đoạn lặp dài (LTR) gồm 15 000 đến 110 000 bp chiếm 8% DNA bộ gen với khoảng 450 000 bản sao.



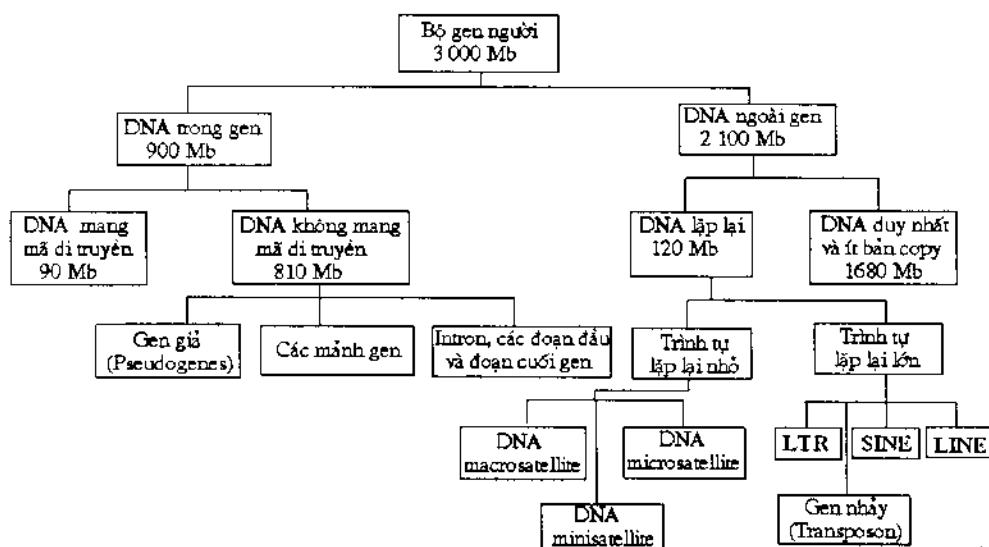
Hình 1.11. Sự phân bố các trình tự Alu trong họ gen globin ở người

Bộ gen người có các họ gen mã hoá protein (họ gen globin, họ gen actin...) và các họ gen mã hoá tRNA, rRNA. Trong bộ gen người có 497 gen mã hoá tRNA phân bố thành các họ gen, trong đó có một họ gen lớn nằm trên nhiễm sắc thể số 6 có kích thước 4 Mb, gồm 140 gen mã hoá tRNA. Các gen mã hoá rRNA tập hợp trong các họ gen nằm ở những vùng DNA lặp lại ít. Gen mã hoá rRNA 28S và rRNA 5, 8S có từ 150 - 200 bản sao, gen mã hoá rRNA 5S gồm 200 - 3000 gen.

Trong bộ gen người ngoài các gen mã hoá protein, các trình tự DNA không mang mã di truyền còn có các gen giả (pseudogene). Gen giả là những gen có cấu trúc giống như các gen mã hoá protein nhưng không được phiên mã, hoặc được phiên mã nhưng không được dịch mã.

Hệ gen tì thể người là phân tử DNA xoắn kép, dạng vòng kích thước khoảng 16569 bp chứa 44% (G + C), gồm 37 gen, trong đó 24 gen mã hoá tRNA và 13 gen mã hoá 13 chuỗi polypeptid. Mã di truyền hệ gen tì thể người có một số khác biệt so với mã di truyền của các gen nhân. Một số đột biến nhỏ trong hệ gen tì

thể, gây nên các bệnh ở người như hội chứng liệt thần kinh thính giác di truyền Leber (LHON), hội chứng bại não Leigh (Leigh syndrome), chứng lười vận động, chứng dư thừa acid lactic trong tế bào. Một số nghiên cứu mới đây của các nhà khoa học Trung Quốc cho rằng đột biến trong hệ gen ti thể có liên quan đến bệnh tiểu đường Typ II...

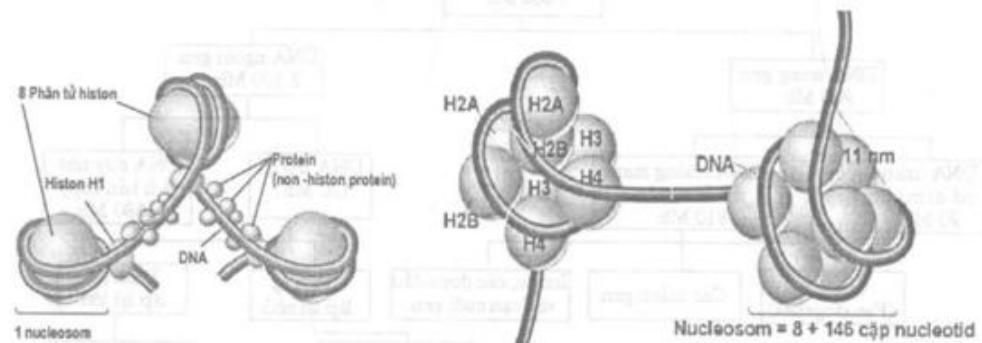


*Hình 1.12. Sự phân bố các nhóm DNA trong bộ gen người*

### 2.3.3. Cấu trúc nhiễm sắc thể

Nhiễm sắc thể của sinh vật *eukaryote* cấu trúc từ các nucleosom cuộn xoắn. Nucleosom được cấu tạo do phân tử DNA cuộn xoắn 1,75 vòng quanh phân lõi histon, mỗi vòng xoắn DNA gồm 80 cặp nucleotid. Phân lõi histon gồm 8 phân tử histon, gồm 2 phân tử histon H<sub>2</sub>A (khối lượng phân tử khoảng 14 kD), 2 phân tử histon H<sub>2</sub>B (khoảng 14 kD), 2 phân tử histon H<sub>3</sub> (khoảng 15 kD), 2 phân tử histon H<sub>4</sub> (khối lượng khoảng 11 kD).

Mỗi nucleosom có đường kính 110 Å ( $1 \text{ Å} = 10^{-7} \text{ mm}$ ), các nucleosom nối với nhau bởi một đoạn DNA kích thước 50 - 70 bp và một phân tử histon H<sub>1</sub> tạo thành chuỗi polynucleosom gọi là sợi cơ bản có đường kính 110 Å. Sợi cơ bản cuộn xoắn zigzag tạo thành sợi nhiễm sắc, đường kính khoảng 300 Å, các sợi nhiễm sắc xoắn gấp khúc khi đạt cực đại tạo thành các nhiễm sắc tử khoảng 700 Å, một nhiễm sắc thể điển hình gồm hai nhiễm sắc tử có đường kính khoảng 1400 - 1500 Å. Một số ít loài sinh vật bậc cao có nhiễm sắc thể rất lớn, đường kính nhiễm sắc thể có thể tới 7 000 Å.



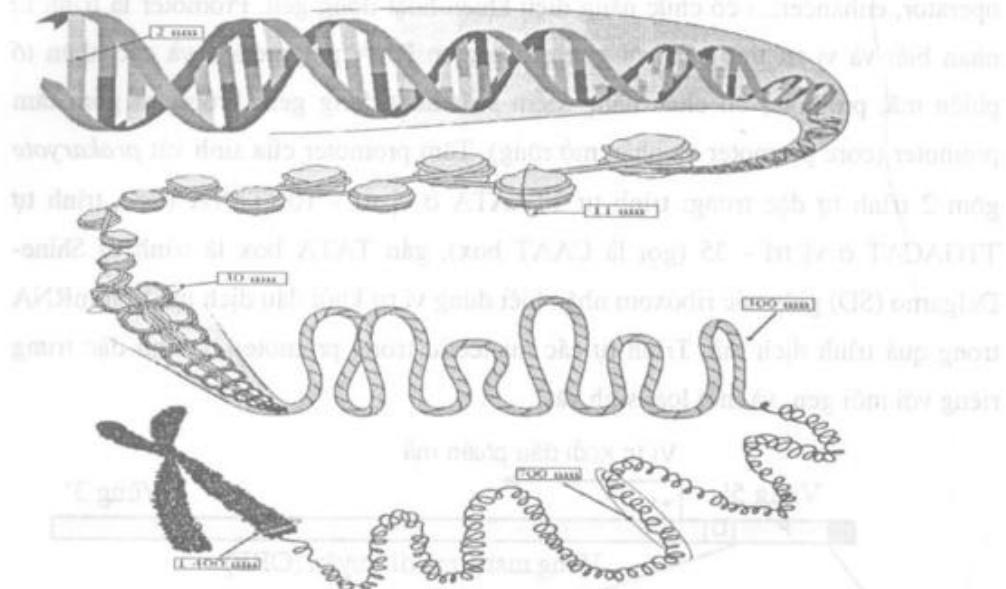
Hình I. 13. Cấu trúc phân tử của nucleosom

Cấu trúc cơ bản của một nhiễm sắc thể gồm tâm động, cánh dài (q), cánh ngắn (p) và điểm mút (telomer). Tâm động là điểm bám của thoi dây tơ vô sắc, giúp cho sự phân li của các nhiễm sắc thể về các cực tế bào khi tế bào phân chia. Telomer là vùng DNA đầu mứt của nhiễm sắc thể giữ vai trò đảm bảo cấu trúc ổn định của nhiễm sắc thể. Telomer của các nhiễm sắc thể ở người cấu trúc từ sự lặp lại hàng trăm lần trình tự 5'-TTAGGG-3'.

Nhiều nghiên cứu gần đây cho rằng, độ dài của telomer liên quan đến tuổi thọ của cá thể, do độ dài telomer góp phần quyết định số lần phân chia tế bào và thời gian tồn tại dài hay ngắn của một tế bào.

Một số nghiên cứu cấu trúc telomer tế bào bách cát người cho thấy kích thước telomer có sự biến động theo lứa tuổi: ở trẻ mới sinh telomer có kích thước trung bình 8 000 bp, ở lứa tuổi 35 độ dài khoảng 3 000 bp, người ở lứa tuổi 65 có

kích thước telomer khoảng 1 500 bp. Khẳng định độ dài telomer liên quan đến tuổi thọ của người, một số nghiên cứu của các nhà khoa học thuộc Đại học Utah (Mỹ) cho thấy những người sống trên 100 tuổi, có kích thước telomer dài hơn rất nhiều so với người chết dưới 65 tuổi.



Hình 1.14. Cấu trúc phân tử của nhiễm sắc thể

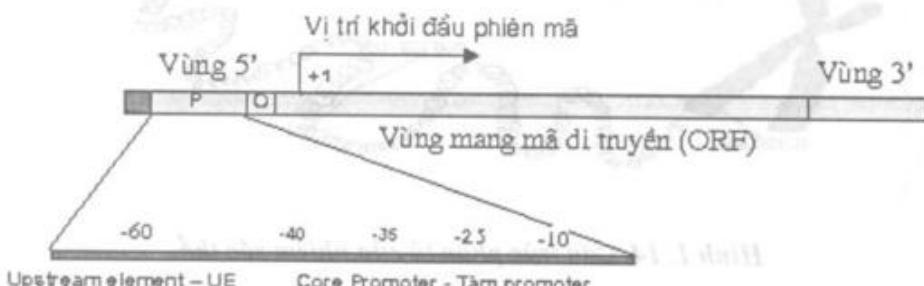
#### IV. Cấu trúc gen của prokaryote và eukaryote

##### 1. Cấu trúc gen mã hoá protein

Bộ gen của sinh vật gồm các trình tự DNA trong gen và trình tự DNA ngoài gen. DNA trong bộ gen gồm 2 nhóm gen chủ yếu là các gen mang mã di truyền và gen không mang mã di truyền. Gen không mang mã di truyền gồm các đoạn DNA lặp lại nhiều, các gen giả (pseudogene), gen đệm hay mảnh gen (spacer gen), các intron...

Gen mang mã di truyền hay còn gọi là gen diễn hình là các đoạn DNA hoặc RNA, mang thông tin di truyền xác định cấu trúc của một chuỗi polypeptid hoặc một phân tử RNA nhất định. Một gen mang mã di truyền có cấu trúc diễn hình gồm ba vùng chính:

Vùng điều khiển hay vùng 5' có một số trình tự đặc hiệu (promoter, operator, enhancer...) có chức năng điều khiển hoạt động gen. Promoter là trình tự nhận biết và vị trí tiếp cận với gen của enzym RNA polymerase và các nhân tố phiên mã, promoter có chức năng kiểm soát hoạt động gen. Promoter gồm tâm promoter (core promoter và phần mở rộng). Tâm promoter của sinh vật *prokaryote* gồm 2 trình tự đặc trưng: trình tự TTAATA ở vị trí -10 (TATA box), trình tự TTGACAT ở vị trí -35 (gọi là CAAT box), gần TATA box là trình tự Shine-Dalgarno (SD) giúp các riboxom nhận biết đúng vị trí khởi đầu dịch mã trên mRNA trong quá trình dịch mã. Trình tự các nucleotid trong promoter có tính đặc trưng riêng với mỗi gen, và mỗi loài sinh vật.

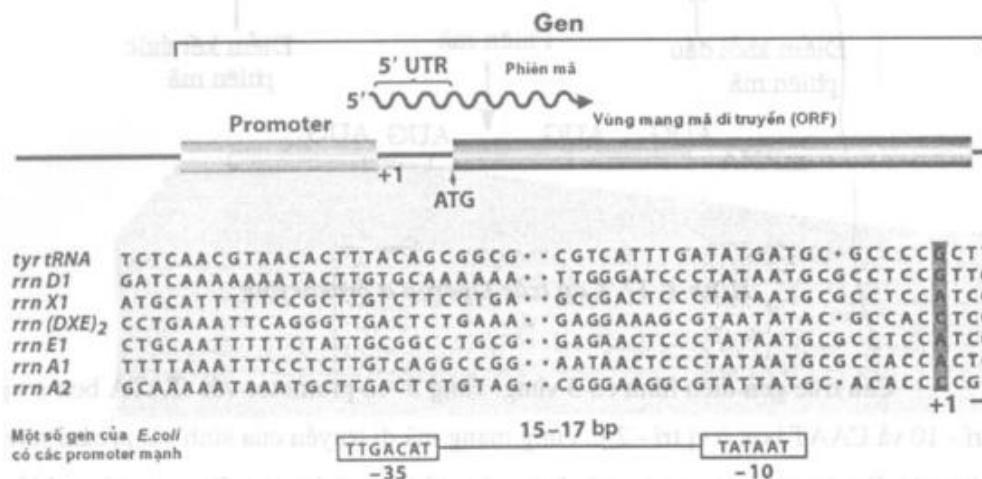


Hình 1.15. Sơ đồ cấu trúc promoter và các vùng gen

Vùng mang mã di truyền (coding sequence) hay khung đọc mã mở (ORF - open reading frame), được phiên mã tạo thành các RNA thông tin (mRNA), hoặc các loại RNA khác (tRNA, rRNA, small RNA...). Các mRNA có thể được dịch mã tạo nên các sản phẩm protein.

Vùng kết thúc hay vùng 3' gồm các trình tự tín hiệu kết thúc, tín hiệu gắn đuôi poly A ở *eukaryote*, trình tự kết thúc (terminator) và một số trình tự lặp ngắn như rRNA, tRNA...

Cấu trúc gen mang mã di truyền còn có một số trình tự đặc trưng khác như các trình tự diêu hoà, trình tự tăng cường (enhance), trình tự bất hoạt (silencer) và các trình tự giữa gen (intergenic)...

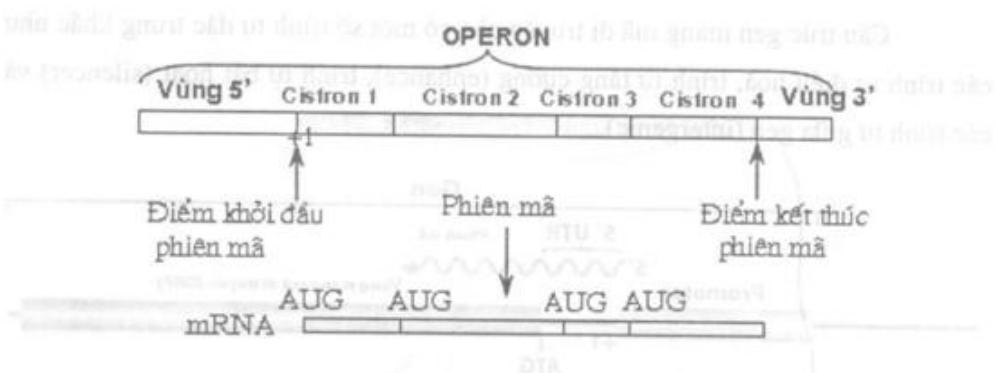


Hình 1. 16. Cấu trúc một số promoter mạnh ở các gen của *E. coli*

Enhancer và silcencer là các trình tự có vai trò điều hoà diều hoà hoạt động gen, enhancer và silcencer có thể nằm ở vị trí trong vùng promoter, nằm trong vùng mang mã di truyền của gen hoặc nằm trong các đoạn intergenic ở phía trước hoặc phía sau gen. Hoạt động của các trình tự enhancer và silcencer làm thay đổi cường độ hoạt động của gen, hoặc làm ngừng hoạt động phiến mã của gen. Enhancer có thể làm tăng cường mức độ tái bản hoặc phiến mã của gen lên hàng trăm, hàng nghìn lần. Ngược lại, trình tự silencer làm giảm cường độ phiến mã hoặc gây ngừng hoạt động của gen.

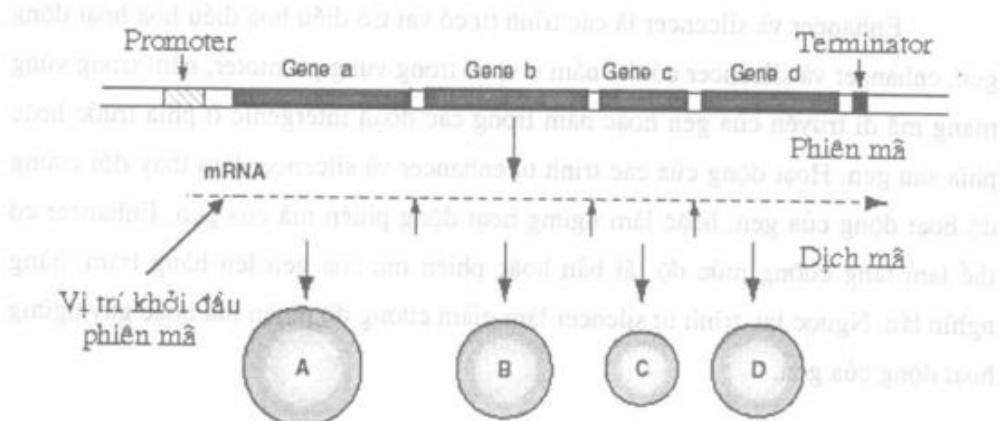
## 2. Đặc điểm cấu trúc gen mã hoá protein của prokaryote

Trong bộ gen của prokaryote các gen có cấu trúc tương đối đồng nhất, phần lớn các gen mang mã di truyền, nhiều gen cùng hướng trao đổi chất tập hợp thành các operon.



Hình 1.17. Cấu trúc Operon ở prokaryote

Cấu trúc gen điển hình có 3 vùng: vùng 5' có promoter với TATA box ở vị trí - 10 và CAAT box ở vị trí - 25. Vùng mang mã di truyền của sinh vật *prokaryote* chủ yếu là các trình tự mang mã di truyền, không có intron. Operon gồm nhiều cistron có chung vùng điều khiển và vùng kết thúc, mỗi cistron tương đương với một gen mã hoá protein ở sinh vật bậc cao.



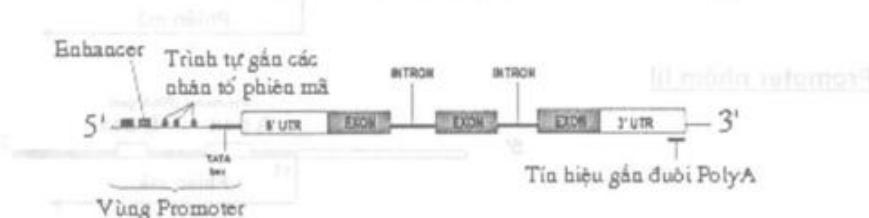
Hình 1.18. Sơ đồ phiên mã và dịch mã ở một Operon ở prokaryote

Trong một operon, quá trình phiên mã khởi đầu từ cistron thứ nhất đến cistron cuối cùng không có quá trình chặn mã và có thể có mã đồng

thời với dịch mã. Hiện tượng các gen cùng hướng trao đổi chất tập hợp thành các operon, tương đối phổ biến ở vi khuẩn, giúp cho vi khuẩn thích ứng nhanh với điều kiện môi trường.

### 3. Đặc điểm cấu trúc gen mã hoá protein của eukaryote

Bộ gen của các sinh vật *eukaryote* gồm các gen có vùng điều khiển và vùng kết thúc riêng, không có hiện tượng tạo thành operon. Các gen mã hoá protein ở sinh vật *eukaryote* phần lớn là gen phân đoạn, mỗi gen gồm các đoạn DNA mang mã di truyền (exon) xen kẽ với các đoạn DNA không mang mã di truyền (intron). Các loài sinh vật ở các bậc càng cao trong hệ thống tiến hoá, có hàm lượng DNA trong các intron của mỗi gen càng lớn.



Hình 1.19. Cấu trúc chung một gen mang mã di truyền của eukaryote

Các gen của sinh vật *prokaryote* có một loại promoter cấu trúc tương đối giống nhau. Các gen mang mã di truyền của sinh vật *eukaryote* được chia làm 3 nhóm chủ yếu với 3 loại promoter khác nhau.

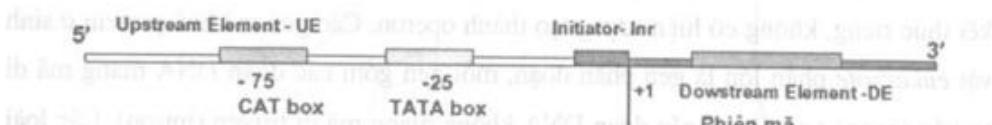
Promoter nhóm I là promoter của các gen mã hoá các loại RNA ribosom (rRNA 18S, rRNA 28S và rRNA 5, 8S). Promoter nhóm I có 2 trình tự đặc trưng: tâm promoter (core promoter) nằm ở vị trí - 40 đến + 20 và trình tự UCE (upstream control element) ở vị trí - 156 đến - 107 trong vùng 5' UTR của gen (Robert Tjian và cộng sự, 2002).

Promoter nhóm II là promoter của các gen mã hoá protein và một số gen mã hoá các loại RNA kích thước nhỏ (small RNA U<sub>1</sub>, U<sub>2</sub>, U<sub>3</sub>...). Cấu trúc promoter nhóm

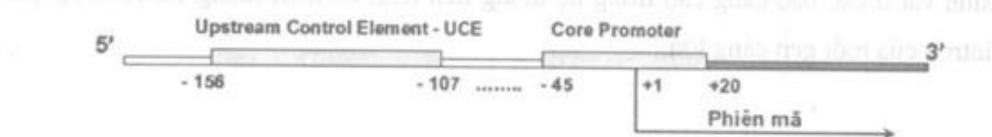
II gồm 4 thành phần: tâm promoter, trình tự UP (upstream element), trình tự khởi đầu (Inr) và trình tự DE (Downstream element).

Promoter nhóm III là promoter của các gen mã hóa các loại tRNA, rRNA 5S và một số phân tử RNA kích thước nhỏ (small RNA như U<sub>5</sub>, U<sub>6</sub>...), trình tự đặc trưng của các promoter nhóm III đang được tiếp tục nghiên cứu.

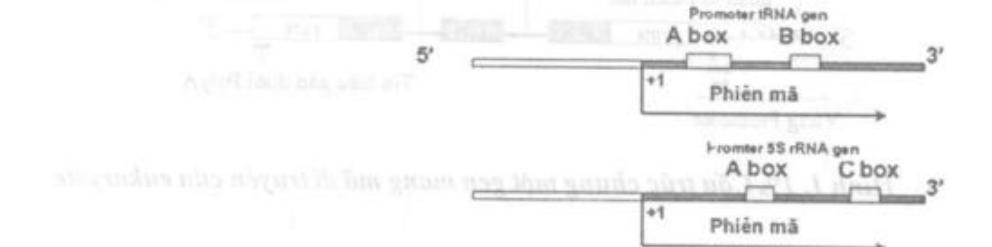
### Promoter nhóm II



### Promoter nhóm I



### Promoter nhóm III

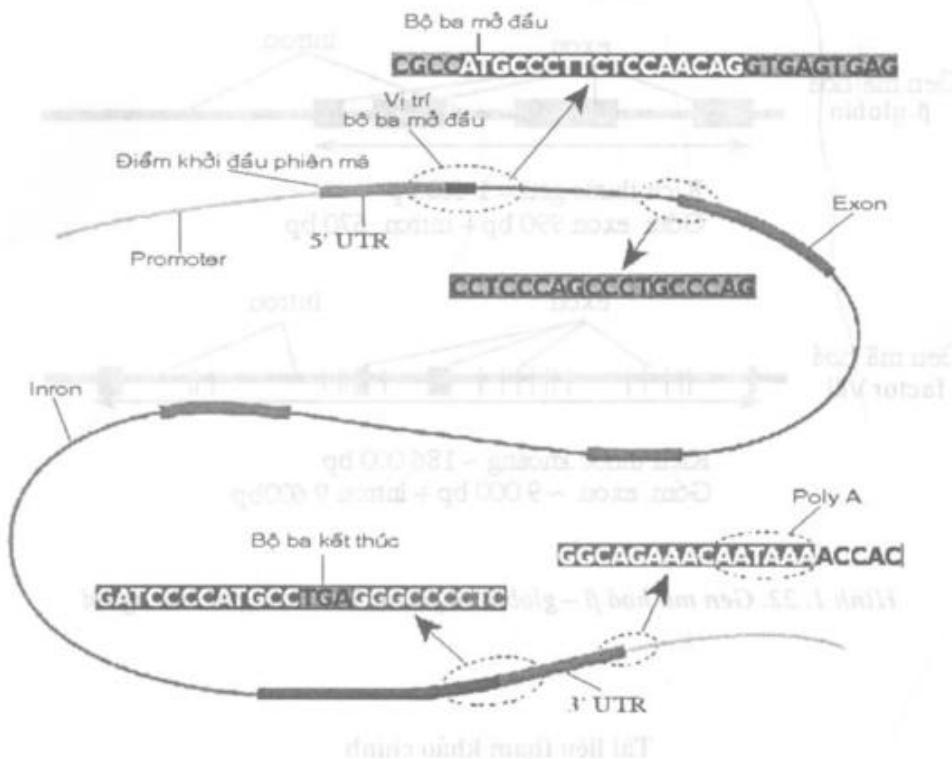


Hình 1.20. Sơ đồ cấu trúc các loại promoter ở sinh vật eukaryote

Vùng mang mã di truyền của sinh vật eukaryote có cấu trúc phức tạp, mỗi gen mã hóa protein mang thông tin di truyền mã hóa một chuỗi polypeptid xác định. Phần lớn gen mã hóa protein gồm các đoạn mang mã di truyền (exon), xen kẽ với các intron là các đoạn không mang mã di truyền (no coding sequences).

Kích thước một intron có thể từ vài chục bazơ nitơ đến hàng nghìn bazơ nitơ, khác nhau tùy từng gen ở loài sinh vật. Ví dụ, các intron của gen ovalbumin trong lòng trắng trứng gà chiếm 75% chiều dài gen; gen mã hóa

$\beta$ -globin người có intron chiếm gần 25% kích thước gen, gen mã hoá yếu tố VIII ở người có intron chiếm tới gần 95% hàm lượng DNA trong gen.

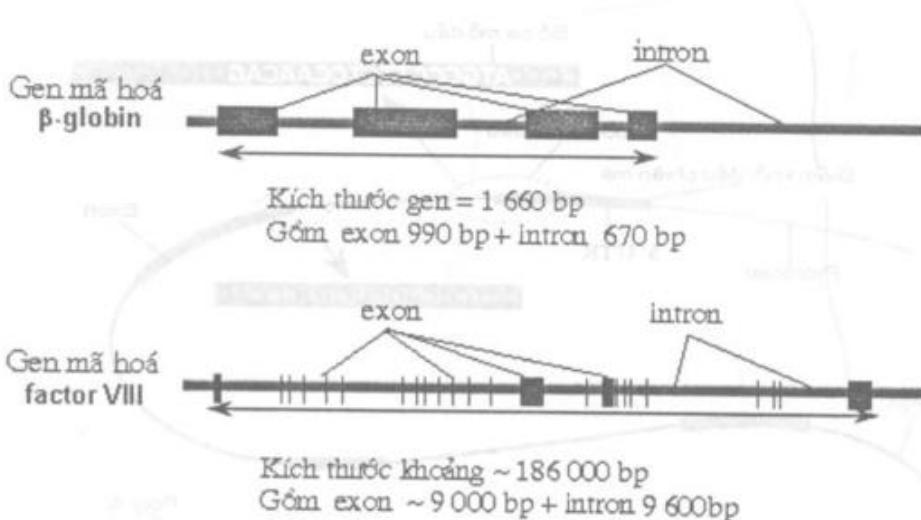


Hình 1.21. Sơ đồ các trình tự đặc hiệu của gen mã hóa protein

Trong quá trình phiên mã các exon và intron đều được phiên mã tạo thành phân tử tiền mRNA (pre mRNA), phân tử tiền mRNA phải qua một quá trình cắt nối (processing) loại bỏ các intron thành mRNA trưởng thành trong nhân tế bào. Sau đó các phân tử mRNA trưởng thành qua màng nhân ra tế bào chất, tham gia quá trình dịch mã thành protein.

Đến nay, vai trò của các intron trong gen chưa được biết đầy đủ, các đột biến trong intron cũng làm thay đổi cấu trúc protein được dịch mã. Ví dụ, bệnh máu khó đông (Hemophilia) ở người trong nhiều trường hợp bị bệnh có liên quan đến các đột biến ở intron 18 và intron 19. Intron có thể có vai trò quan trọng trong sự

biểu hiện thông tin di truyền ở eukaryote, hoặc chức năng ổn định của cấu trúc gen và cấu trúc nhiễm sắc thể.



Hình 1. 22. Gen mã hoá  $\beta$  – globin và factor VIII trong tế bào người

#### Tài liệu tham khảo chính

1. Brown T. A., (2002). Genomes. BIOS Scientific, Oxford, United Kingdom.
2. Bruford M. W., Wayne R. K., (1993). Microsatellites and their application to population genetic studies. *Genet. Dev.* 3: 939-943.
3. Dale J. W., Schantz M., (2002). From Genes to Genomes: Concepts and Applications of DNA Technology. John Wiley & Sons, Ltd.
4. Goldman N., (2001). Molecular phylogenetics: State-of-the-art methods for looking into the past, *Trends in Genetics* 17, 262 – 272.
5. Hartwell L. H., Hood L., Goldberg M. L., Reynolds A.E., Silver L.S., and Veres R. C., (2004). Genetics: From genes to genomes, 2nd edition. McGraw Hill, New York.

6. Hartl D. L., Jones E. W., (1999). Essential Genetics. Bartlett Pub., USA
7. Khuất Hữu Thanh, (2003). Cơ sở di truyền phân tử và kỹ thuật gen. NXB Khoa học và Kỹ thuật, Hà Nội.
8. Watson J. D. và cộng sự (2004), Molecular Biology of the Gene. Fifth edition, Pearson Education, Inc., pub. Benjamin Cummings.
9. Weaver R. F., (2002). Molecular Biology, Mc Graw Hill.
10. Weaver R. F., Hedrick P.W., (1997). Genetics. Brown Pub., USA.
11. Wu S. B., Collins G., Sedgley M. (2004), “A molecular linkage map of olive (*Olea europaea* L) based on RAPD, microsatellite, and SCAR markers”, *Genome*, 47(1), pp. 26-35.
12. <http://evolution.genetics.washington.edu/phylip/software.html>
13. <http://www.mips.biochem.mpg.de/proj/yeast/catalogues/funcat/>)
14. <http://bioweb.pasteur.fr/GenoList/SubtiList>
15. <http://genome-www.stanford.edu/Saccharomyces>
16. <http://www.mips.biochem.mpg.de/proj/yeast>
17. <http://bmb.med.miami.edu/ecogene>

## **Chương 2**

### **BẢN ĐỒ DI TRUYỀN**

#### **I. Vài nét về lược sử nghiên cứu bản đồ di truyền**

Khi nghiên cứu cấu trúc di truyền của một loài sinh vật, một cá thể, một nhiễm sắc thể hoặc một gen... người ta thu được các dữ liệu di truyền. Sự biểu hiện các dữ liệu di truyền bằng hình dạng, kích thước, màu sắc của nhiễm sắc thể hoặc bằng độ lớn, vị trí sắp xếp của các gen, các nucleotid... tạo nên các loại bản đồ di truyền khác nhau. Bản đồ di truyền cho phép xác định mức độ quan hệ, hoặc so sánh sự khác nhau về mặt di truyền giữa các đối tượng sinh vật. Mặt khác, bản đồ di truyền cho phép dự đoán một số tính chất, đặc điểm quý hiếm của sinh vật hoặc chẩn đoán sớm một số bệnh di truyền, bệnh truyền nhiễm của cá thể sinh vật...

Tuỳ theo phương pháp nghiên cứu, mục đích, đối tượng nghiên cứu và hình thức thể hiện, người ta chia bản đồ di truyền thành nhiều loại khác nhau: bản đồ di truyền liên kết, bản đồ di truyền tế bào, bản đồ vật lí, bản đồ trình tự gen...

Năm 1910, nhà di truyền học người Mỹ T. Morgan bằng các kết quả nghiên cứu đặc điểm di truyền của ruồi giấm (*Drosophila melanogaster*), đã xây dựng bản đồ liên kết gen đầu tiên. Năm 1931, C. Stern phát hiện cơ sở tế bào học của tái tổ hợp ở ruồi giấm, nhà di truyền thực vật McClintock phát hiện cơ chế tái tổ hợp và các yếu tố di truyền vận động ở ngô..., một số bản đồ di truyền liên kết gen đầu tiên ở ruồi giấm và ở cây ngô được xây dựng.

Năm 1934, với các phương pháp nhuộm màu đặc hiệu các nhà khoa học đã phát hiện nhiễm sắc thể khổng lồ của ruồi giấm. Sau đó, bằng phương pháp nhuộm giemsa và quan sát trên kính hiển vi, các nhà khoa học đã phát hiện nhiễm sắc thể khổng lồ của ruồi đầm, có khoảng hơn 5 000 vạch màu đậm nhạt khác nhau và đã

xây dựng bản đồ di truyền tế bào (cytogenetic map) nhiễm sắc thể ruồi giấm. Những năm gần đây, ứng dụng các phương pháp nhuộm huỳnh quang, nhiều bản đồ di truyền tế bào của các loài sinh vật khác nhau được xây dựng.

Năm 1957, Wallman và Jacob sử dụng kỹ thuật giao nạp (tiếp hợp) ngắt quãng xây dựng được nên bản đồ gen giao nạp của vi khuẩn *E. coli*.

Năm 1970, Hamilton Smith phát hiện enzym giới hạn (RE - Restriction Enzyme) là các enzym có khả năng cắt đặc hiệu phân tử DNA ở những trình tự nhất định, nhiều loại enzym giới hạn khác nhau được sử dụng để nghiên cứu bộ gen và xây dựng bản đồ giới hạn (Restriction map) của các loài sinh vật khác nhau.

Năm 1977, Alan Maxam và Walter Gilbert phát minh phương pháp xác định trình tự gen bằng phương pháp hoá học. Đồng thời, Frederick Sanger và cộng sự phát minh phương pháp giải trình tự gen theo phương pháp dideoxy, bản đồ trình tự bộ gen đầu tiên của thực khuẩn thể φ X174 được Sanger và cộng sự xây dựng đầu tiên năm 1977.

Cùng với sự phát triển của khoa học công nghệ, nhiều phương pháp nghiên cứu mới được sử dụng trong nghiên cứu bộ gen của các loài sinh vật, các loại bản đồ di truyền mới như bản đồ lai phóng xạ, bản đồ lai tại chỗ huỳnh quang (FISH), bản đồ tạo dòng định vị (positional cloning)... được xây dựng.

Bản đồ trình tự gen (sequence map) là loại bản đồ có độ chính xác cao, được sử dụng rộng rãi hiện nay trong các nghiên cứu bộ gen của virus, vi khuẩn và các loài sinh vật bậc cao. Bản đồ trình tự gen có thể xác định chính xác vị trí từng nucleotid trong bộ gen. Bản đồ trình tự bộ gen góp phần xác định nguồn gốc phân loại, sự tiến hoá của các quần thể hoặc loài sinh vật, đồng thời giúp con người xác định được những gen liên quan đến các tính trạng quý hiếm ở vật nuôi cây trồng, hoặc các gen đột biến, gen bị sai lệch dẫn đến các bệnh do rối loạn di truyền ở người...

Bộ gen người được giải trình tự hoàn chỉnh năm 2003, bộ gen cây lúa được giải trình tự đầy đủ năm 2005 đánh dấu một bước tiến vĩ đại trong công nghệ sinh học phân tử và kỹ thuật gen. Nghiên cứu bộ gen người giúp các nhà khoa học có cơ sở nghiên cứu nguồn gốc tiến hoá của loài người, xác định nguyên nhân một số

bệnh di truyền, bệnh truyền nhiễm ở người và các sinh vật khác. Từ kết quả giải mã bộ gen người, con người có cơ sở khoa học để phát triển các loại thuốc chữa bệnh mới, các liệu pháp chữa bệnh hiệu quả cao, góp phần bảo vệ sức khoẻ và nâng cao đời sống con người.

## **II. Một số loại bản đồ di truyền chủ yếu**

Trong nghiên cứu di truyền học, với các mục đích và phương pháp nghiên cứu khác nhau, người ta đã xây dựng được các loại bản đồ di truyền khác nhau của các loài sinh vật nghiên cứu. Mỗi loại bản đồ di truyền có vai trò và tính chất đặc trưng riêng.

Tuỳ theo mục đích nghiên cứu, sự tiến bộ của khoa học công nghệ và thời gian lịch sử, người ta có thể xây dựng, sử dụng các loại bản đồ di truyền khác nhau. Chẳng hạn, trong nửa đầu thế kỉ 20, sử dụng nhiều nhất các loại bản đồ hình thái nhiễm sắc thể và bản đồ di truyền liên kết. Ngày nay, bản đồ trình tự gen được sử dụng nhiều hơn cả trong nghiên cứu gen và bộ gen các loài sinh vật.

Một số loại bản đồ di truyền được nghiên cứu và sử dụng chủ yếu gồm:

Bản đồ hình thái nhiễm sắc thể hay bản đồ kiểu nhân (Karyotype map)

Bản đồ di truyền liên kết (Genetic linkage)

Bản đồ di truyền giao nạp (Conjugation map)

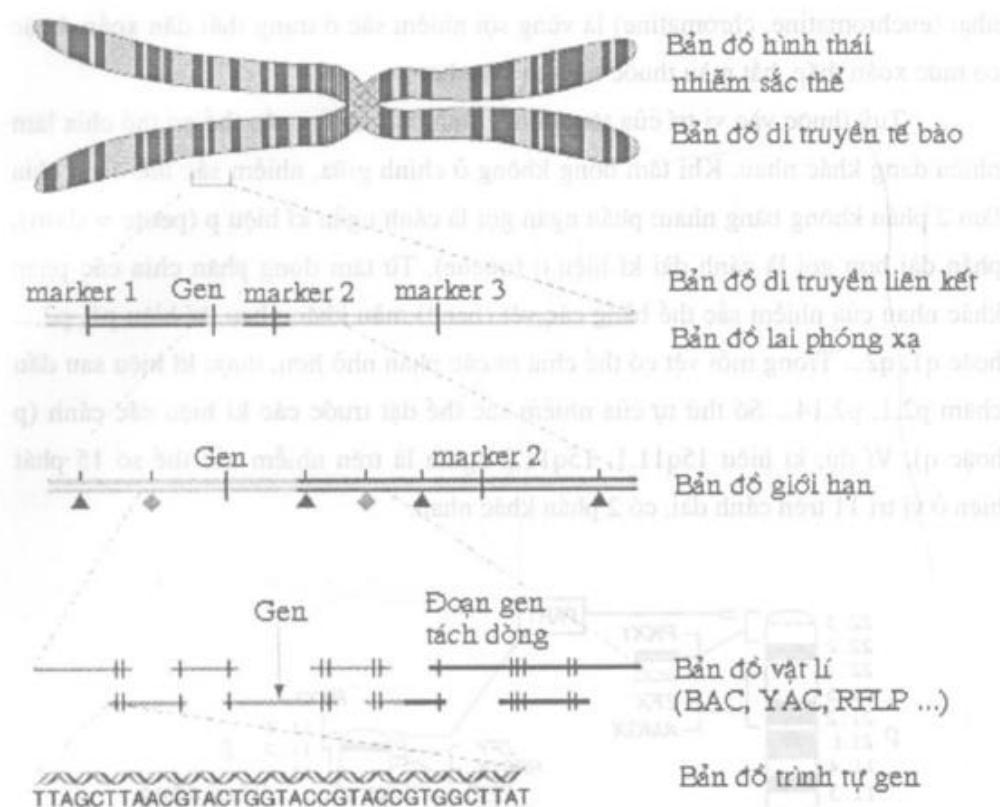
Bản đồ di truyền tế bào (Cytogenetic map)

Bản đồ lai phóng xạ (Radiantion Hybrides map)

Bản đồ di truyền giới hạn (Restriction map)

Bản đồ vật lí (Physical mapping): là tên gọi chung của nhiều loại bản đồ di truyền được lập bằng các phương pháp khác nhau: bản đồ BAC, YAC, STS, EST, VNTR, RFLP...

Bản đồ trình tự gen (Sequence map)...



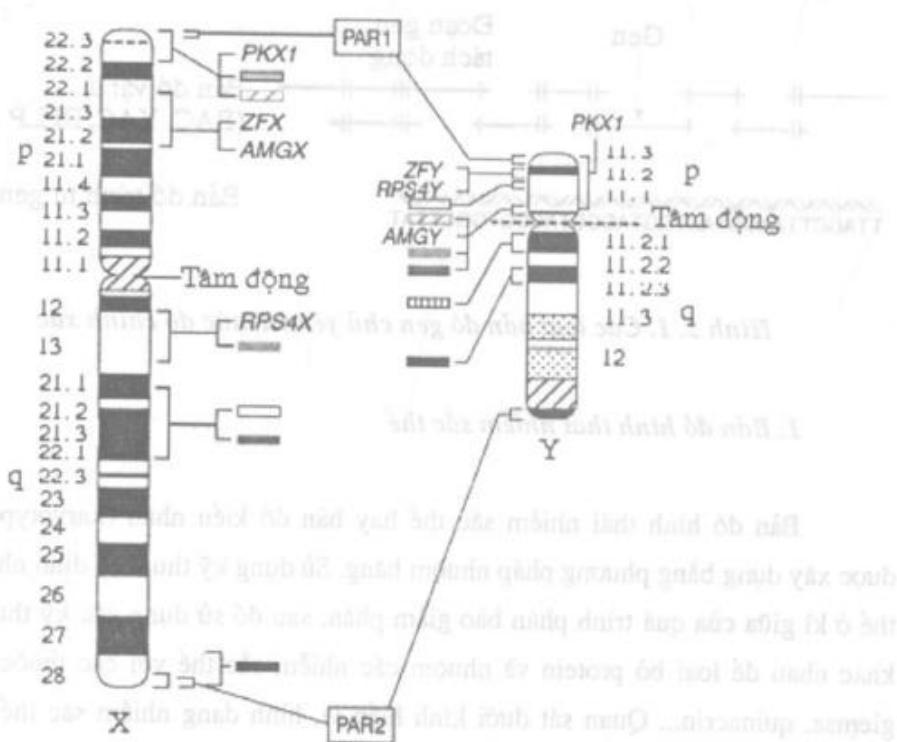
Hình 2.1. Các loại bản đồ gen chủ yếu và mức độ chính xác

### 1. Bản đồ hình thái nhiễm sắc thể

Bản đồ hình thái nhiễm sắc thể hay bản đồ kiểu nhân (karyotype map), được xây dựng bằng phương pháp nhuộm băng. Sử dụng kỹ thuật cố định nhiễm sắc thể ở kì giữa của quá trình phân bào giảm phân, sau đó sử dụng các kỹ thuật xử lí khác nhau để loại bỏ protein và nhuộm các nhiễm sắc thể với các thuốc nhuộm giemsa, quinacrin... Quan sát dưới kính hiển vi, hình dạng nhiễm sắc thể và tần động biểu hiện ở các vệt (band) màu đậm nhạt khác nhau trong cấu trúc nhiễm sắc thể. Vùng dị nhiễm sắc (heterochromatine) là vùng sợi nhiễm sắc có mức độ xoắn cao, bắt màu mạnh với các thuốc nhuộm, tạo nên các vệt màu đậm. Vết nhuộm màu

nhạt (euchromatine, chromatine) là vùng sợi nhiễm sắc ở trạng thái dãn xoắn, hoặc có mức xoắn thấp, bắt màu thuốc nhuộm kém hơn.

Tùy thuộc vào vị trí của tâm động, hình thái nhiễm sắc thể có thể chia làm nhiều dạng khác nhau. Khi tâm động không ở chính giữa, nhiễm sắc thể được chia làm 2 phần không bằng nhau: phần ngắn gọi là cánh ngắn kí hiệu p (petite = short), phần dài hơn gọi là cánh dài kí hiệu q (queue). Từ tâm động phân chia các phần khác nhau của nhiễm sắc thể bằng các vệt (band) màu khác nhau, kí hiệu p1, p2,... hoặc q1, q2... Trong mỗi vệt có thể chia ra các phần nhỏ hơn, được kí hiệu sau dấu chấm p2.1, p2.14... Số thứ tự của nhiễm sắc thể đặt trước các kí hiệu các cánh (p hoặc q). Ví dụ, kí hiệu 15q11.1, 15q11.2 nghĩa là trên nhiễm sắc thể số 15 phát hiện ở vị trí 11 trên cánh dài, có 2 phần khác nhau.



**Hình 2.2. Bản đồ hình thái nhiễm sắc thể X và Y ở người**

Hiện nay có nhiều phương pháp nhuộm băng khác nhau. Diễn hình là các phương pháp nhuộm băng G, nhuộm băng R, Q, C,...

Phương pháp nhuộm băng G: cố định nhiễm sắc thể và xử lí để loại bỏ protein bằng NaOH, sau đó nhuộm giemsa. Trên nhiễm sắc thể hiện lên các vạch màu đậm nhạt khác nhau. Các băng đậm thu được là vùng DNA giàu A - T, có mức cuộn xoắn cao, còn các băng nhạt là vùng DNA giàu G - C, có độ cuộn xoắn thấp.

Phương pháp nhuộm băng R: cố định nhiễm sắc thể và gây biến tính DNA bằng nhiệt với các dung dịch đệm thích hợp, sau đó nhuộm giemsa. Kết quả ngược lại với nhuộm băng G, các băng đậm màu là vùng DNA giàu G - C, các băng nhạt màu là vùng giàu A - T.

Phương pháp nhuộm băng Q: cố định nhiễm sắc thể và nhuộm bằng quinacrin, các băng đậm màu hiện lên là vùng giàu A - T, các băng nhạt màu là vùng giàu G - C.

Phương pháp nhuộm băng C: biến tính DNA trên nhiễm sắc thể bằng bari hydroxid bão hòa, sau đó nhuộm giemsa. Kết quả các băng màu sẫm là vùng di nhiễm sắc (heterochromatin), DNA ở các vệt này có độ lặp lại nhiều và mức độ xoắn rất cao.

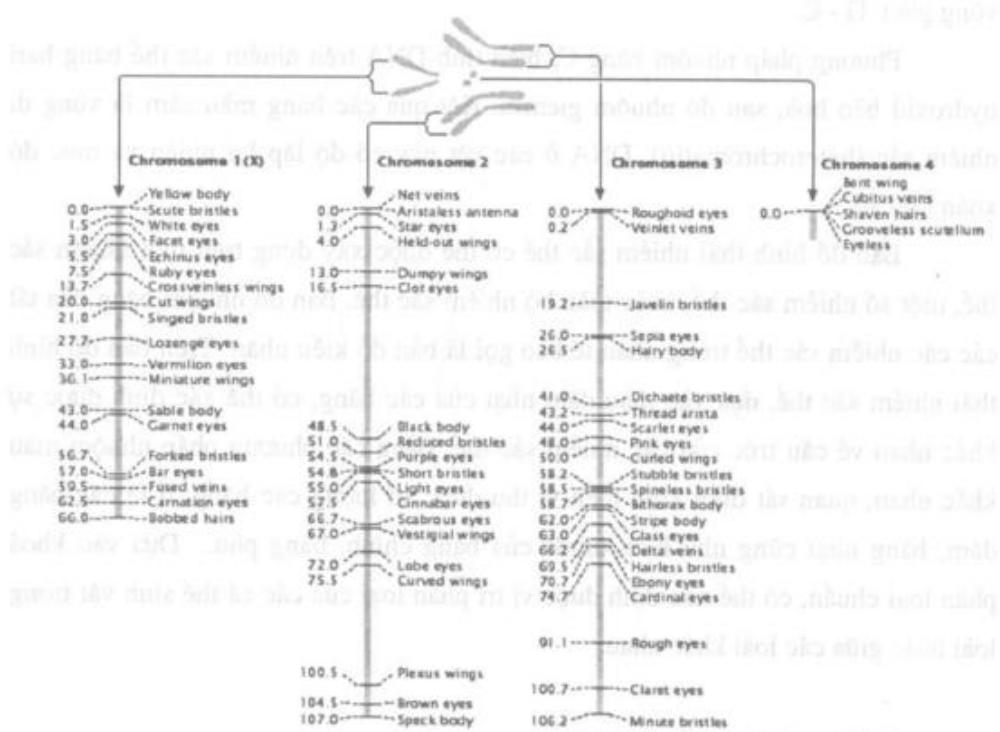
Bản đồ hình thái nhiễm sắc thể có thể được xây dựng trên một nhiễm sắc thể, một số nhiễm sắc thể, hoặc toàn bộ nhiễm sắc thể. Bản đồ nhuộm băng của tất cả các nhiễm sắc thể trong nhân tế bào gọi là bản đồ kiểu nhân. Trên bản đồ hình thái nhiễm sắc thể, dựa vào màu đậm nhạt của các băng, có thể xác định được sự khác nhau về cấu trúc của các nhiễm sắc thể. Bằng các phương pháp nhuộm màu khác nhau, quan sát dưới kính hiển vi thu được số lượng các băng, tỉ lệ các băng đậm, băng nhạt cũng như kích thước của băng chính, băng phụ... Dựa vào khoa phân loại chuẩn, có thể xác định được vị trí phân loại của các cá thể sinh vật trong loài hoặc giữa các loài khác nhau.

## 2. Bản đồ di truyền liên kết

Bộ gen sinh vật bậc cao có số lượng gen rất lớn (lúa có hơn 37 000 gen,

muỗi 13 683 gen, ở người có khoảng 32 000 gen...). Số lượng nhiễm sắc thể của mỗi loài rất ít (lúa  $2n = 24$ , người  $2n = 46...$ ), do đó trên mỗi nhiễm sắc thể gồm rất nhiều gen, các gen liên kết với nhau trên nhiễm sắc thể tạo thành các nhóm liên kết gen. Số nhóm liên kết gen bằng số nhiễm sắc thể đơn bội trong bộ nhiễm sắc thể. Ví dụ, ở lúa có số nhóm liên kết gen là  $n = 12$ ; ruồi giấm có số nhóm liên kết gen  $n = 4$ .

Bản đồ di truyền truyền liên kết (GL - Genetic linkage, genetic map) được xây dựng dựa trên tần số tái tổ hợp giữa các locus gen trên các cặp nhiễm sắc thể tương đồng. Bản đồ di truyền liên kết xây dựng đầu tiên vào năm 1910 - 1913, từ kết quả các nghiên cứu trên ruồi giấm của nhà di truyền học người Mỹ T. Morgan. Tần số tái tổ hợp giữa 2 cặp gen là tần số trao đổi chéo và hoán vị gen, xảy ra giữa các locus gen của các cặp nhiễm sắc thể tương đồng trong giảm phân.

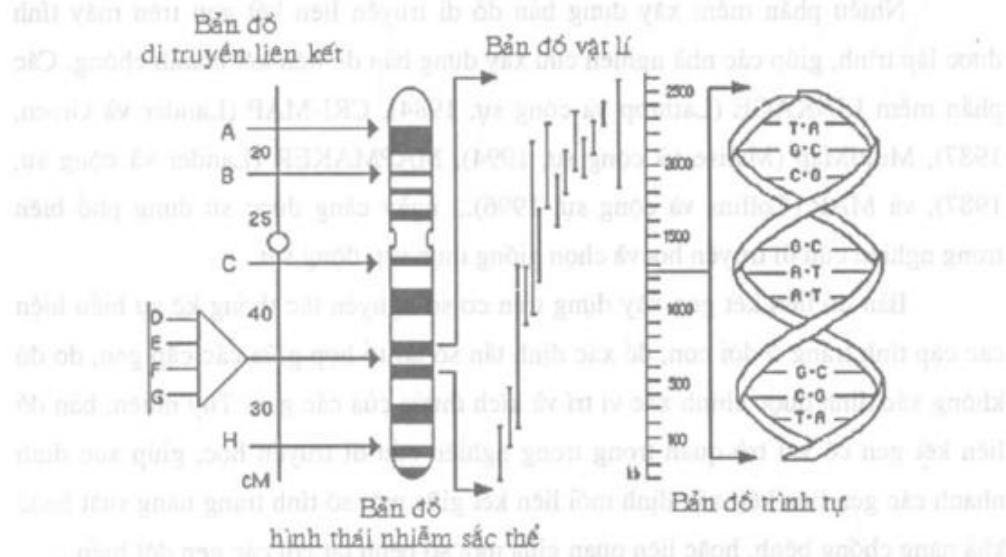


Hình 2.3. Bản đồ liên kết gen ở ruồi giấm

Tần số tái tổ hợp gen được xác định bằng tỉ lệ các cá thể tái tổ hợp/ tổng số các thế thu được trong quần thể đời con. Từ tần số tái tổ hợp giữa các gen, xác định được khoảng cách tương đối giữa các gen trên nhiễm sắc thể, tính bằng đơn vị centimorgan (cM).

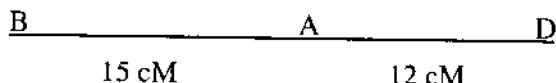
Trên bản đồ liên kết gen, khoảng cách giữa các gen trên nhiễm sắc thể tính bằng centimorgan (cM). Qui ước, 1 cM = 1% tần số trao đổi chéo (tần số tái tổ hợp) giữa 2 cặp gen. Nhiều công trình nghiên cứu sau này cho rằng 1cM tương đương với 1 Mb ở các loại bản đồ di truyền khác.

Bản đồ di truyền liên kết chỉ xác định được vị trí tương đối giữa các gen trên nhiễm sắc thể. Tần số tái tổ hợp giữa 2 cặp gen ( $f\%$ ) biến động từ 0 - 50%, khi  $f = 0\%$  nghĩa là trong quần thể không có tái tổ hợp giữa các cặp gen nghiên cứu. Theo lý thuyết, tần số tái tổ hợp cao nhất bằng 50%, nhưng trong thực tế tần số tái tổ hợp giữa 2 cặp gen thường rất nhỏ. Bản đồ di truyền liên kết xây dựng trên cơ sở tần số tái tổ hợp giữa 2 locus gen, mà tần số tái tổ hợp gen không chỉ xảy ra một cách ngẫu nhiên giữa hai các gen, mà còn xảy ra với các cặp gen khác trên nhiễm sắc thể. Do đó, vị trí giữa các gen trên bản đồ di truyền liên kết không hoàn toàn tương ứng với vị trí các gen xác lập được bằng các loại bản đồ vật lí.



Hình 2.4. Mối liên quan giữa các loại bản đồ gen

Trong mỗi nhóm liên kết số lượng gen trên thường rất lớn, tái tổ hợp gen không chỉ xảy ra giữa 2 locus gen mà còn xảy ra đồng thời với nhiều locus gen trên nhiễm sắc thể, nên thiết lập bản đồ di truyền liên kết vô cùng phức tạp. Để thiết lập bản đồ liên kết gen, cần phải xác định được tần số tái tổ hợp giữa từng cặp gen với nhau, từ đó suy ra vị trí tương đối giữa các gen trên nhiễm sắc thể. Ví dụ, trên một nhiễm sắc thể (cùng một nhóm liên kết) xét 3 locus gen A, B và D. Kết quả thực nghiệm thu được tần số tái tổ hợp giữa A - D = 12%, A - B = 15%, B - D = 27% , từ đó suy ra vị trí các gen trên nhiễm sắc thể là:



Ngày nay, ngoài tần số tái tổ hợp giữa các gen, các dữ liệu đa hình di truyền của các trình tự đích (STS - sequence targeted site), tính đa hình của các đoạn lặp DNA vi vê tinh (microsatellite) hoặc tính đa hình của các nucleotid đơn (SNP – single nucleotide polymorphism)...cũng được sử dụng để xây dựng bản đồ di truyền liên kết gen, làm tăng tính chính xác của bản đồ liên kết gen.

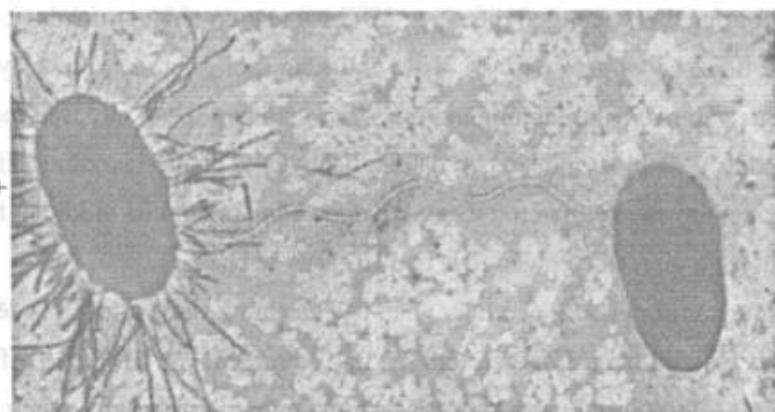
Nhiều phần mềm xây dựng bản đồ di truyền liên kết gen trên máy tính được lập trình, giúp các nhà nghiên cứu xây dựng bản đồ liên kết nhanh chóng. Các phần mềm LINKAGE (Lathrop và cộng sự, 1984), CRI-MAP (Lander và Green, 1987), MultiMap (Matise và cộng sự, 1994), MAPMAKER (Lander và cộng sự, 1987), và MAP (Collins và cộng sự, 1996)... ngày càng được sử dụng phổ biến trong nghiên cứu di truyền học và chọn giống thực vật, động vật.

Bản đồ liên kết gen xây dựng trên cơ sở nguyên tắc thống kê sự biểu hiện các cặp tính trạng ở đời con, để xác định tần số tái tổ hợp giữa các cặp gen, do đó không xác định được chính xác vị trí và kích thước của các gen. Tuy nhiên, bản đồ liên kết gen có vai trò quan trọng trong nghiên cứu di truyền học, giúp xác định nhanh các gen liên kết, xác định mối liên kết giữa một số tính trạng năng suất hoặc khả năng chống bệnh, hoặc liên quan giữa một số bệnh tật với các gen đột biến...

**3. Bản đồ di truyền giao nạp** là bản đồ di truyền bộ gen ở vi khuẩn, được xây dựng trên cơ sở phương pháp giao nạp ngắt quãng. Năm 1953, William Hayes phát hiện nhân tố giới tính ( $F$ ) ở *E. coli*. Tế bào *E. coli* có nhân tố giới tính  $F$  gọi là tế bào  $F^+$ , tế bào không có nhân tố giới tính gọi là tế bào  $F^-$ . Các tế bào *E. coli*  $F^+$  có thể giao nạp (conjugation) với tế bào  $F^-$  qua ống giao nạp.

Năm 1957, Wallman và F. Jacob, sử dụng kỹ thuật giao nạp ngắt quãng (làm đứt ống giao nạp bằng máy rung) ở từng thời điểm nhất định, đã lập được bản đồ bộ gen của *E. coli*.

Năm 1957, Wallman và F. Jacob, sử dụng kỹ thuật giao nạp ngắt quãng (làm đứt ống giao nạp bằng máy rung) ở từng thời điểm nhất định, đã lập được bản đồ bộ gen của *E. coli*.



**Hình 2.5. Giao nạp ở vi khuẩn *E. coli***

Vi khuẩn *E. coli* có sự tái tổ hợp giữa plasmid giới tính với bộ gen, làm cho plasmid giới tính gắn với bộ gen gọi là episom. Chủng vi khuẩn *E. coli* có episom có khả năng tái tổ hợp cao gọi là chủng Hfr (high frequency recombination). Khi tế bào Hfr giao nạp với tế bào *E. coli*  $F^-$ , tế bào Hfr có thể chuyển nhân tố giới tính  $F$ , và một phần DNA của tế bào sang tế bào vi khuẩn thể nhận  $F^-$  qua ống giao nạp.

Wallman và Jacob cho các tế bào dòng Hfr bình thường giao nạp với các tế bào *E. coli* F<sup>-</sup> mang 7 gen đột biến, để nghiên cứu quá trình giao nạp và thiết lập bản đồ giao nạp ở vi khuẩn *E. coli*.

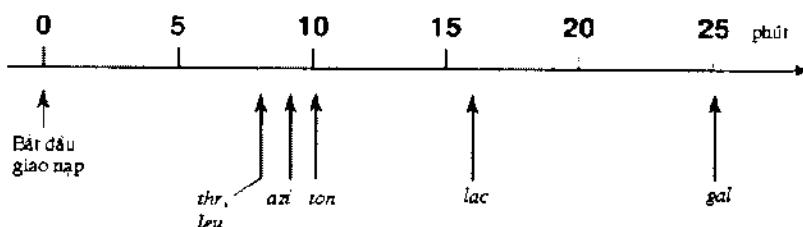
Các gen đột biến (*thr<sup>-</sup>*) và (*leu<sup>-</sup>*) làm mất khả năng tổng hợp acid amin threonin và leucin; đột biến (*azi<sup>S</sup>*) mẫn cảm với độ acid của môi trường; đột biến (*tonA<sup>S</sup>*) mẫn cảm với phage T<sub>1</sub>; đột biến (*lac<sup>-</sup>*) và (*gal<sup>-</sup>*) làm mất khả năng thuỷ phân đường lactoza và đường galactoza; đột biến (*Str<sup>R</sup>*) chịu được chất kháng sinh streptomycin.

Dòng Hfr bình thường: *thr<sup>+</sup>, leu<sup>+</sup>, azi<sup>R</sup>, tonA<sup>R</sup>, lac<sup>+</sup>, gal<sup>+</sup>* và *str<sup>R</sup>*

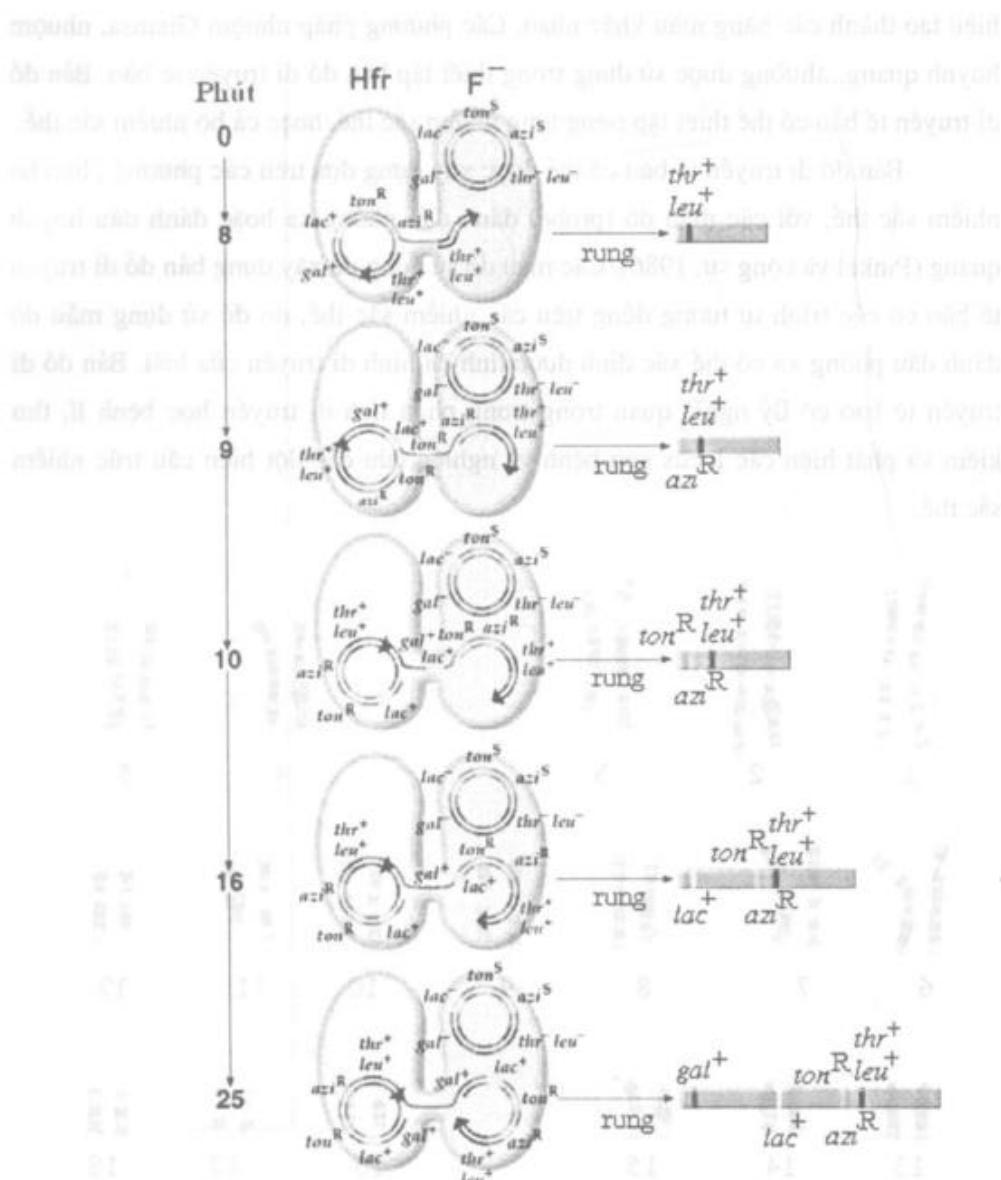
Dòng F<sup>-</sup> đột biến: *thr<sup>-</sup>, leu<sup>-</sup>, azi<sup>S</sup>, tonA<sup>S</sup>, lac<sup>-</sup>, gal<sup>-</sup>* và *str<sup>S</sup>*

Nuôi các tế bào *E. coli* dòng Hfr cùng với tế bào F<sup>-</sup> trong các ống nghiệm trong môi trường dinh dưỡng, ở điều kiện 37°C với mật độ tế bào thích hợp, để các tế bào có thể tiếp xúc và giao nạp với nhau. Sau từng thời gian nhất định lấy các ống nghiệm, đưa vào máy rung làm ngừng quá trình giao nạp. Nuôi cấy trên môi trường chọn lọc có 2 % để kiểm tra các gen đã được chuyển từ các tế bào dòng Hfr sang tế bào F<sup>-</sup>. Tổng hợp các kết quả thu được, lập được bản đồ gen của vi khuẩn tính bằng đơn vị phút (thời gian chuyển gen từ tế bào Hfr sang tế bào F<sup>-</sup>).

Kết quả thí nghiệm, cho thấy nhân tố giới tính F được chuyển từ tế bào Hfr sang tế bào F<sup>-</sup> sớm nhất. Các gen *thr<sup>+</sup>* và *leu<sup>+</sup>* được chuyển từ tế bào Hfr sang tế bào F<sup>-</sup> sau 6 - 7 phút giao nạp, gen *azi<sup>R</sup>* không mẫn cảm với acid được chuyển từ tế bào Hfr sang F<sup>-</sup> sau 8 phút giao nạp, gen *tonA<sup>R</sup>* không mẫn cảm với phage T<sub>1</sub> được chuyển từ tế bào Hfr sang F<sup>-</sup> sau 10 phút, gen *lac<sup>+</sup>* sau 17 phút, gen *gal<sup>+</sup>* sau 23 phút. Thời gian cần thiết chuyển hết DNA của Hfr sang F<sup>-</sup> mất 90 phút.



**Hình 2. 6. Bản đồ giao nạp ở vi khuẩn *E. coli***



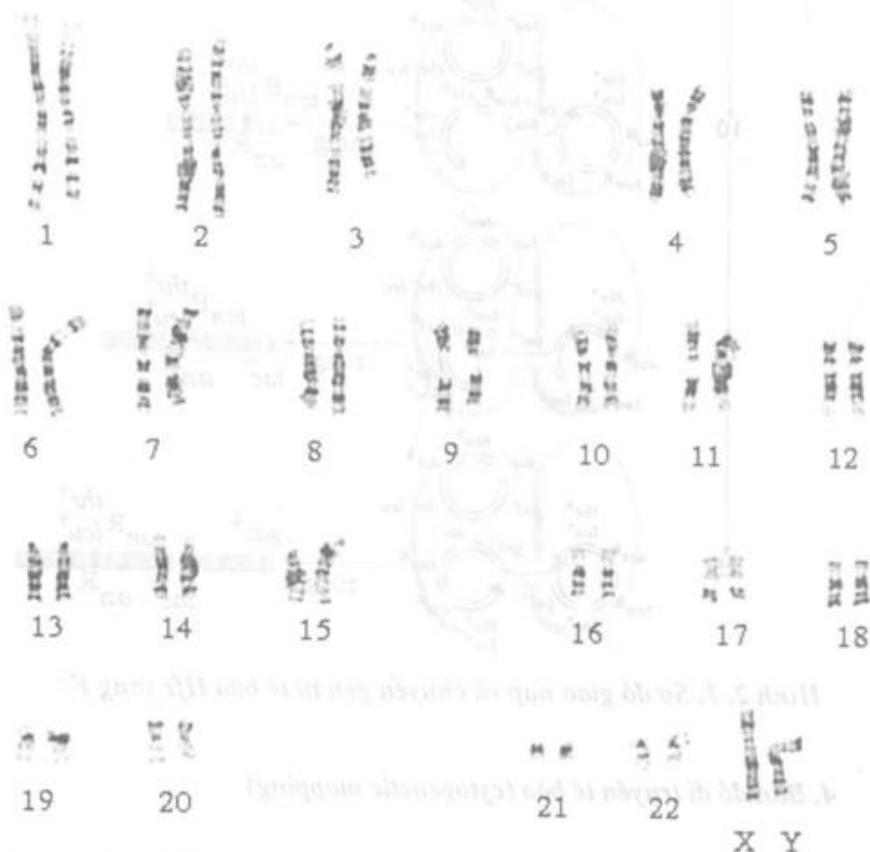
Hình 2.7. Sơ đồ giao nạp và chuyển gen từ tế bào Hfr sang F<sup>-</sup>

#### 4. Bản đồ di truyền tế bào (cytogenetic mapping)

Bản đồ di truyền tế bào xây dựng trên cơ sở các phân khác nhau của nhiễm

hiệu tạo thành các băng màu khác nhau. Các phương pháp nhuộm Giemsa, nhuộm huỳnh quang...thường được sử dụng trong thiết lập bản đồ di truyền tế bào. Bản đồ di truyền tế bào có thể thiết lập riêng từng nhiễm sắc thể, hoặc cả bộ nhiễm sắc thể.

Bản đồ di truyền tế bào có thể được xây dựng dựa trên các phương pháp lai nhiễm sắc thể, với các mẫu dò (probe) đánh dấu phóng xạ hoặc đánh dấu huỳnh quang (Pinkel và cộng sự, 1986). Các mẫu dò sử dụng để xây dựng bản đồ di truyền tế bào có các trình tự tương đồng trên các nhiễm sắc thể, do đó sử dụng mẫu dò đánh dấu phóng xạ có thể xác định được tính đa hình di truyền của loài. Bản đồ di truyền tế bào có ý nghĩa quan trọng trong phân tích di truyền học bệnh lí, tìm kiếm và phát hiện các locus gen bệnh và nghiên cứu các đột biến cấu trúc nhiễm sắc thể.



Hình 2.8. Bản đồ di truyền tế bào người ( $2n = 46$ )

Bản đồ di truyền tế bào phân biệt tương đối rõ rệt các vùng nhiễm sắc thể với kích thước khoảng 10 Mb, các đoạn nhiễm sắc thể kích thước nhỏ (dưới 10 Mb) khó xác định chính xác bằng bản đồ di truyền tế bào.

Hiện nay, với kỹ thuật lai huỳnh quang tại chỗ FISH (Fluorescence In Situ Hybridization, Lawrence và cộng sự, 1990) có thể xây dựng được các bản đồ FISH, nâng cao độ chính xác của bản đồ di truyền tế bào. Bản đồ di truyền FISH, có thể xác định và phân biệt các đoạn nhiễm sắc thể kích thước từ 1- 2 Mb với độ chính xác cao. Phương pháp FISH cho phép xác định nhanh các gen đột biến, gen gây bệnh trên nhiễm sắc thể, giúp con người đưa ra giải pháp điều trị hữu hiệu nhất. Tuy nhiên, phương pháp FISH không áp dụng được với những locus gen có kích thước lớn.

### *5. Bản đồ lai phóng xạ*

Bản đồ lai phóng xạ (RH map - radiation hybrid mapping) là bản đồ di truyền được xây dựng trên cơ sở khuếch đại các đoạn nhiễm sắc thể bị đứt gãy do xử lý phóng xạ, được lai với các nhiễm sắc thể bình thường tạo nên các dạng thể khảm nhiễm sắc thể. Kỹ thuật xử lý phóng xạ được phát hiện năm 1970, khi chiếu xạ tế bào người bằng tia X với liều cao 3 000 - 8 000 röntgen, gây đứt gãy các nhiễm sắc thể thành các đoạn kích thước khác nhau.

Kỹ thuật xây dựng bản đồ lai phóng xạ được hoàn thiện năm 1990 (Cox và cộng sự, 1990). Nhiều loài sinh vật đã được xây dựng bản đồ lai phóng xạ (Stewart và cộng sự, 1997), bản đồ lai phóng xạ bộ gen người được xây dựng năm 1996 gồm 40322 đoạn STS (Olivier và cộng sự, 2001).

Xử lý tế bào với liều phóng xạ cao, làm cho nhiễm sắc thể bị đứt gãy thành nhiều đoạn nhỏ, còn xử lý tế bào ở liều phóng xạ thấp có thể gây đứt gãy nhiễm sắc thể thành một số đoạn kích thước lớn. Ví dụ, Stewart (1997) sử dụng liều phóng xạ cao 10 000 rad xử lý nhiễm sắc thể, tạo nên các đoạn đứt gãy nhiễm sắc thể có kích thước trung bình 4 Mb.

**Bảng 2. 1. Liều xử lí phóng xạ và kích thước đoạn nhiễm sắc thể bị đứt gãy**

	GeneBridge4 (GB4)	Stanford (G3)
Liều phóng xạ	3 000 rad	10 000 rad
Kích thước đoạn đứt gãy	10 Mb	4 Mb
Mức độ đứt gãy	ít	trung bình
Tác giả	Gyapay và cs ,1996	Stewart và cs ,1997

Thiết lập bản đồ lai phóng xạ cần thực hiện một số bước cơ bản sau:

- Cố định tế bào cần nghiên cứu ở giai đoạn kì giữa của quá trình phân bào, xử lí tia X với liều phóng xạ cao làm đứt nhiễm sắc thể thành nhiều đoạn nhỏ. Căn cứ vào mục đích nghiên cứu, và đặc điểm cấu trúc nhiễm sắc thể của mỗi loài sinh vật lựa chọn liều phóng xạ, để tạo được các đoạn nhiễm sắc thể có kích thước phù hợp.

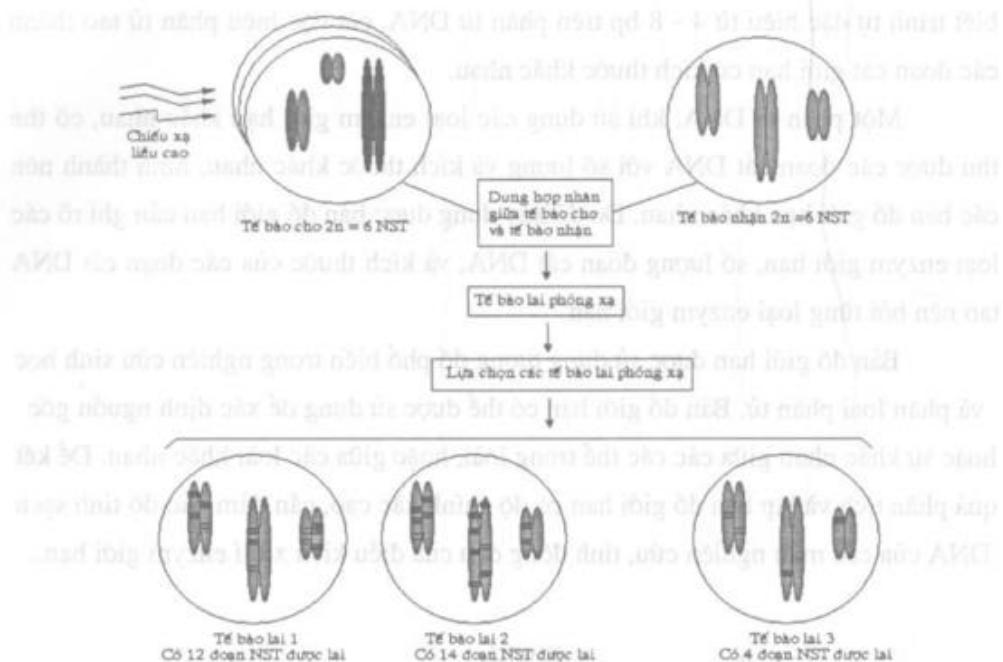
1- Thực hiện kỹ thuật lai tế bào đã xử lí phóng xạ với một tế bào bình thường khác nhờ kỹ thuật PEG (Polyethylen Glycol) hoặc nhờ virus Sendai... làm dung hợp nhân giữa hai tế bào. Ví dụ, lai tế bào người đã xử lí phóng xạ với tế bào chuột Hamster.

2- Nuôi cấy tế bào lai trong môi trường dinh dưỡng thích hợp, đảm bảo quá trình dung hợp nhân của tế bào lai, hình thành các tế bào lai có các nhiễm sắc thể ở dạng thể khâm. Ví dụ, khi lai tế bào người xử lí phóng xạ liều cao với tế bào chuột Hamster (CHO) bình thường, tạo nên thể khâm nhiễm sắc thể (các đoạn nhiễm sắc thể người được gắn vào nhiễm sắc thể của chuột Hamster).

3- Sử dụng kỹ thuật PCR để nhận các đoạn DNA lạ được cài gắn trong nhiễm sắc thể khâm của tế bào lai. Tập hợp các kết quả, tính toán tổng số đoạn nhiễm sắc thể bị đứt gãy trong tế bào được xử lí phóng xạ, và số lượng đoạn nhiễm sắc thể được gắn với nhiễm sắc thể của tế bào chủ để xây dựng bản đồ lai phóng xạ.

Đơn vị đo trong bản đồ lai phóng xạ tính bằng centi Ray (cR). Centi Ray tính bằng lần số tương đối giữa số lượng đoạn nhiễm sắc thể lạ được gắn trong tế

bào lai, trên tổng số đoạn nhiễm sắc thể đứt gãy tạo nên do xử lý phóng xạ liều cao. Qui ước: 1 cR bằng 1% số đoạn nhiễm sắc thể đứt giữa 2 locus gen (marker) trong một phép lai đơn.



Hình 2. 9. Sơ đồ các bước lập bản đồ lai phóng xạ

Độ chính xác của bản đồ lai phóng xạ phụ thuộc vào liều lượng phóng xạ, kỹ thuật dung hợp nhân và loại tế bào lai được lựa chọn. Đơn vị do centi Ray (cR) trên bản đồ lai phóng xạ có sự tương quan nhất định với các đơn vị đo của các loại bản đồ di truyền khác. Ví dụ, 1 cR tương đương 2 kb trên bản đồ vật lí.

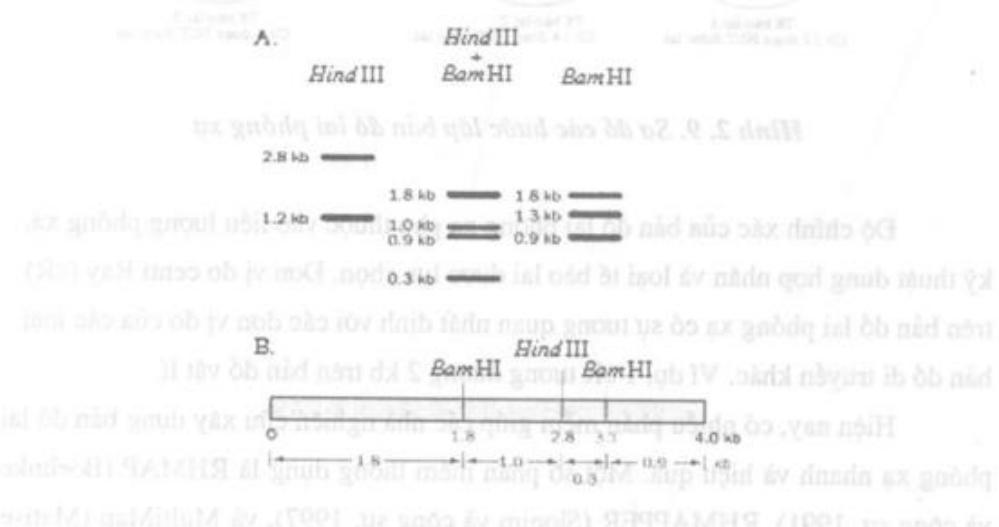
Hiện nay, có nhiều phần mềm giúp các nhà nghiên cứu xây dựng bản đồ lai phóng xạ nhanh và hiệu quả. Một số phần mềm thông dụng là RHMAP (Boehnke và cộng sự, 1991), RHMAPPER (Slonim và cộng sự, 1997), và MultiMap (Matise và cộng sự, 1994)...

**6. Bản đồ giới hạn ( restriction mapping)**

Bản đồ giới hạn là bản đồ di truyền thể hiện số lượng và kích thước của các đoạn DNA bị cắt bằng các loại enzym giới hạn. Enzym giới hạn là các enzym nhận biết trình tự đặc hiệu từ 4 - 8 bp trên phân tử DNA, cắt đặc hiệu phân tử tạo thành các đoạn cắt giới hạn có kích thước khác nhau.

Một phân tử DNA, khi sử dụng các loại enzym giới hạn khác nhau, có thể thu được các đoạn cắt DNA với số lượng và kích thước khác nhau, hình thành nên các bản đồ giới hạn khác nhau. Do đó xây dựng được bản đồ giới hạn cần ghi rõ các loại enzym giới hạn, số lượng đoạn cắt DNA, và kích thước của các đoạn cắt DNA tạo nên bởi từng loại enzym giới hạn.

Bản đồ giới hạn được sử dụng tương đối phổ biến trong nghiên cứu sinh học và phân loại phân tử. Bản đồ giới hạn có thể được sử dụng để xác định nguồn gốc hoặc sự khác nhau giữa các thể trong loài, hoặc giữa các loài khác nhau. Để kết quả phân tích và lập bản đồ giới hạn có độ chính xác cao, cần đảm bảo độ tinh sạch DNA của các mẫu nghiên cứu, tính đồng đều của điều kiện xử lý enzym giới hạn...



Hình 2.10. Bản đồ giới hạn sử dụng enzym BamHI và HindIII

Phương pháp lập bản đồ giới hạn gồm một số bước sau:

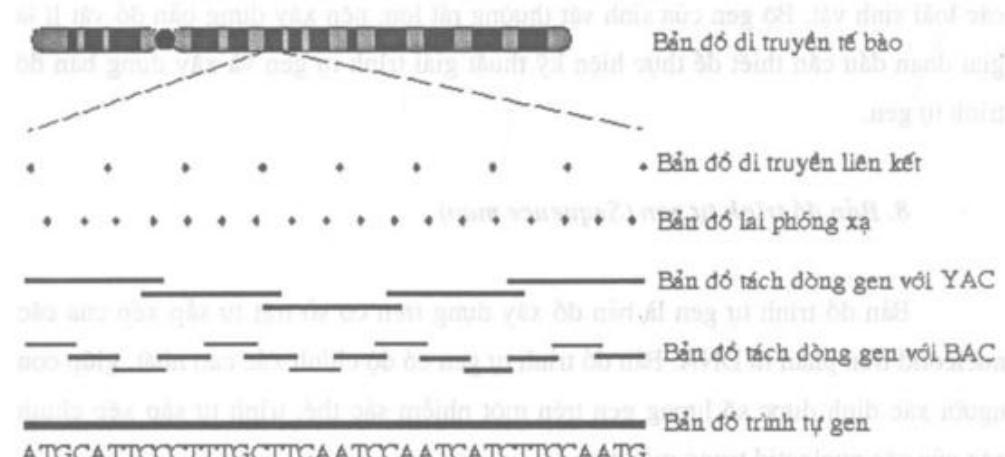
Tách chiết và tinh sạch DNA của mẫu nghiên cứu và thực hiện nhân gen PCR để tạo một lượng DNA cần thiết.

Lựa chọn các enzym giới hạn thích hợp, xử lý enzym giới hạn các mẫu nghiên cứu trong điều kiện thích hợp.

Điện di kết quả trên gel agarose cùng với thang DNA chuẩn, tính toán kết quả và lập bản đồ giới hạn.

### 7. Bản đồ vật lí

Bản đồ vật lí (physical map) là loại bản đồ di truyền được xây dựng từ các đoạn DNA (trình tự DNA) được phân cắt từ bộ gen. Các đoạn DNA sử dụng để xây dựng bản đồ vật lí có thể là các đoạn DNA tách dòng trong các loại vectơ tách dòng khác nhau (BAC, YAC, Cosmid...), các đoạn STS (sequence tagged site) hoặc các đoạn DNA thu được bằng các kỹ thuật phân tử RFLP, SSR..., nên còn gọi là bản đồ tách dòng.



Hình 2.11. So sánh mức độ phân biệt của các loại bản đồ di truyền

Bản đồ vật lí xây dựng dựa vào kích thước của các đoạn DNA, có thể tính bằng số cặp bazơ (bp), kilo bazơ (kb) hoặc Mb ( $1\text{Mb} = 1\,000\text{ kb}$ ,  $1\text{ kb} = 1\,000\text{ bp}$ ). Số lượng các đoạn cắt DNA, kích thước của các đoạn cắt DNA trong mỗi loại bản đồ vật lí mang tính đặc trưng riêng, tạo nên các loại bản đồ vật lí có mức độ chính xác khác nhau. Ví dụ, bản đồ tách dòng với nhiễm sắc thể nhân tạo của vi khuẩn (BAC) có thể phân biệt các đoạn DNA có kích thước nhỏ hơn bản đồ tách dòng với nhiễm sắc thể nhân tạo của nấm men (YAC).

Bộ gen sinh vật bậc cao có kích thước rất lớn, cần được tách dòng trong các vectơ tạo nên các ngân hàng bộ gen khác nhau (YAC, BAC, EST...), sau đó thực hiện lập bản đồ trình tự gen từng phần. Năm 1990, nghiên cứu bộ gen người tách dòng với YAC tạo nên tổng số 33 000 dòng YAC. Mỗi dòng YAC mang một đoạn DNA trong bộ gen người, kích thước khoảng 0,9 Mb (Cohen và cộng sự, 1993). Năm 1995, nghiên cứu bộ gen người các nhà khoa học Đại học Mashachusetts đã tạo được 20 104 STS khi sử dụng kỹ thuật tách dòng các trình tự đích (STS - sequence tagged site).

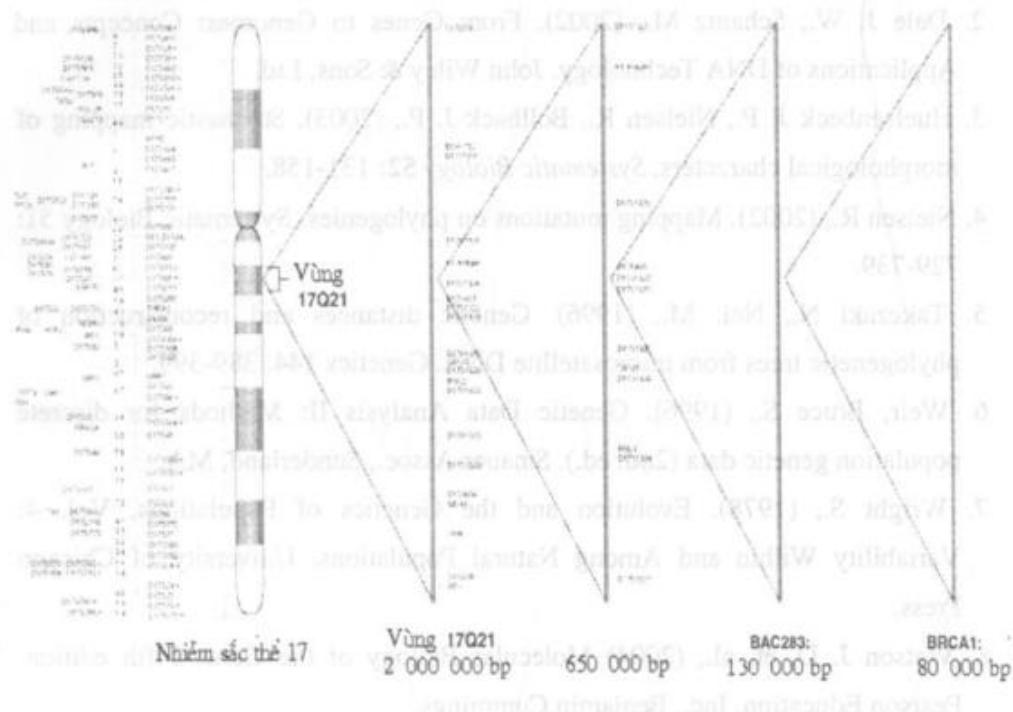
Bản đồ vật lí có ý nghĩa quan trọng giúp lưu giữ thông tin di truyền của cá thể và các loài sinh vật trong các nghiên cứu đa dạng sinh học và vốn gen (genopol) các loài sinh vật. Bộ gen của sinh vật thường rất lớn, nên xây dựng bản đồ vật lí là giai đoạn đầu cần thiết để thực hiện kỹ thuật giải trình tự gen và xây dựng bản đồ trình tự gen.

### **8. Bản đồ trình tự gen (Sequence map)**

Bản đồ trình tự gen là bản đồ xây dựng trên cơ sở trật tự sắp xếp của các nucleotid trên phân tử DNA. Bản đồ trình tự gen có độ chính xác cao nhất, giúp con người xác định được số lượng gen trên một nhiễm sắc thể, trình tự sắp xếp chính xác của các nucleotid trong mỗi gen.

Bản đồ trình tự gen đầu tiên được hoàn thiện là bản đồ gen của thực khuẩn thể φ X 174 ký sinh *E. coli*. Bộ gen phage φ X174 gồm 5 386 nucleotid đã được xác định vị trí sắp xếp của các nucleotid (F. Sanger và cộng sự 1977). Bản đồ trình tự

gen được thiết lập ở nhiều mức độ khác nhau: có thể là trình tự của một gen, các gen trên một nhiễm sắc thể hoặc tất cả bộ gen. Ví dụ, nhiễm sắc thể Y trong bộ gen người được giải mã 1992 (Page và cộng sự), nhiễm sắc thể 21 và nhiễm sắc thể 22 giải mã năm 1999 (Cohen và cộng sự et al ), toàn bộ bộ gen người được giải trình tự 2003, bộ gen cây lúa được giải trình tự hoàn chỉnh năm 2005...



Hình 2. 12. Các bước lập bản đồ trình tự gen ở NST 17 của người

Giải trình tự bộ gen sinh vật bậc cao là quá trình khó khăn, phức tạp đòi hỏi sự cộng tác của nhiều nhà khoa học và chi phí tốn kém cho trang thiết bị. Thiết lập bản đồ trình tự gen gồm một số bước chủ yếu sau: Tách chiết và tinh sạch toàn bộ DNA bộ gen, sử dụng các kỹ thuật phân tử và các loại vectơ tách dòng BAC, YAC,... để tách dòng và xây dựng ngân hàng bộ gen đầy đủ. Giải trình tự từng đoạn DNA tách dòng trong các ngân hàng bộ gen, sau đó tập hợp các dữ liệu để xây dựng bản đồ trình tự gen (phương pháp giải trình tự gen: xem chương 5).

**Tài liệu tham khảo chính**

1. Brown T. A., (2002). Genomes. 2<sup>nd</sup>. ed. Oxford. UK: Bios Scientific Publishers
2. Dale J. W., Schantz M., (2002). From Genes to Genomes: Concepts and Applications of DNA Technology. John Wiley & Sons, Ltd.
3. Huelsenbeck J. P., Nielsen R., Bollback J. P., (2003). Stochastic mapping of morphological characters. *Systematic Biology* 52: 131-158.
4. Nielsen R., (2002). Mapping mutations on phylogenies. *Systematic Biology* 51: 729-739.
5. Takezaki N., Nei. M., (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* 144: 389-399.
6. Weir, Bruce S., (1996). Genetic Data Analysis II: Methods for discrete population genetic data (2nd. ed.). Sinauer Assoc., Sunderland, MA.
7. Wright S., (1978). Evolution and the Genetics of Populations, Vol. 4: Variability Within and Among Natural Populations. University of Chicago Press.
8. Watson J. D. et. al., (2004) Molecular Biology of the Gene. Fifth edition, Pearson Education, Inc., Benjamin Cummings.
9. <http://www.nature.com/genomics/human>
10. <http://www.ncbi.nlm.nih.gov/genome/central>
11. <http://www.celera.com>
12. <http://mbgd.genome.ad.jp>

## Chương 3

### PHÂN TÍCH DỮ LIỆU DI TRUYỀN VÀ PHƯƠNG PHÁP DỤNG CÂY PHÂN LOẠI

#### I. Ý nghĩa của phân tích dữ liệu di truyền và dụng cây phân loại

Sự tiến hoá của các loài sinh vật trong tự nhiên, là kết quả quá trình biến đổi cấu trúc di truyền trong một thời gian lịch sử lâu dài, ở những điều kiện nhất định. Nghiên cứu sự thay đổi cấu trúc DNA, trình tự sắp xếp các nucleotid trong bộ gen hoặc sự thay đổi trình tự acid amin của các phân tử protein của các quần thể, các loài thấy được tiến hoá trong nội bộ loài, giữa các loài hoặc sự khác biệt giữa các cá thể trong quần thể.

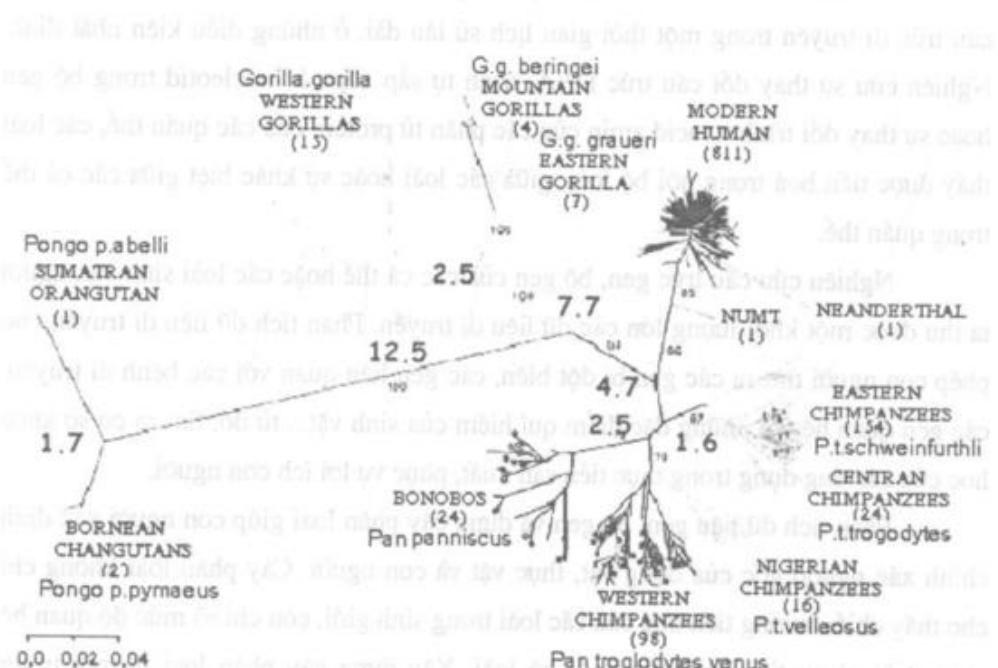
Nghiên cứu cấu trúc gen, bộ gen của các cá thể hoặc các loài sinh vật, người ta thu được một khối lượng lớn các dữ liệu di truyền. Phân tích dữ liệu di truyền cho phép con người tìm ra các gen bị đột biến, các gen liên quan với các bệnh di truyền, các gen quan hệ với những đặc điểm quý hiếm của sinh vật... từ đó, tìm ra cơ sở khoa học của các ứng dụng trong thực tiễn sản xuất, phục vụ lợi ích con người.

Phân tích dữ liệu gen, bộ gen và dụng cây phân loại giúp con người xác định chính xác nguồn gốc của động vật, thực vật và con người. Cây phân loại không chỉ cho thấy chiều hướng tiến hoá của các loài trong sinh giới, còn chỉ rõ mức độ quan hệ trong loài và sự tiến hoá trong nội bộ loài. Xây dựng cây phân loại là một trong những phương pháp phân loại phân tử, có vị trí quan trọng trong các nghiên cứu sự phát sinh của sinh vật trên trái đất, đồng thời xác định chính xác nguồn gốc và sự tiến hoá của các loài sinh vật.

Cây phân loại (phylogeny, dendrogram) được nghiên cứu có hệ thống và khoa học từ những năm đầu thế kỷ 20. Năm 1909, từ các số liệu phân tích tần số các allele trong locus hemoglobin, Reichert và Brown đã xây dựng cây phân loại đầu tiên của một số loài cá.

Năm 1963, Margoliash xây dựng cây phân loại dựa trên trình tự các acid amin trong phân tử protein của cytochrom C (Cyt C). Nhiều công trình nghiên cứu của các tác giả sau đó, trên nhiều đối tượng sinh vật đã đưa ra các phương pháp dựng cây phân loại khác nhau (Hakimi và Yau 1964; Farris 1970; Waterman và cộng sự 1977; Felsenstein 1981; Saitou và Nei 1987; Li 1997)...

Cây phân loại hay cây tiến hoá (Phylogenetic tree, phylogeny, dendrogram) được xây dựng trên cơ sở các dữ liệu di truyền như trình tự nucleotid, số lượng và kích thước các đoạn DNA, trình tự các acid amin trong chuỗi polypeptid hoặc tần số các allele trong locus gen...



Hình 3.1. Cây phân loại chiều hướng tiến hóa từ động vật thành người trên cơ sở dữ liệu phân tích DNA ti thể (Gagneux và cộng sự 1999)

Trong nửa đầu thế kỉ 20, bằng các dữ liệu phân tích tần số allele của các enzym, izoenzym, các tính trạng... người ta đã xây dựng cây phân loại của nhiều loài sinh vật khác nhau, xác định một số quan hệ giữa các loài sinh vật trong tự nhiên.

Hiện nay, từ dữ liệu nghiên cứu trình tự nucleotid trong gen, trình tự acid amin trong phân tử protein con người đã và đang hoàn thiện cây phân loại của nhiều loài virus, vi khuẩn, thực vật, động vật... với độ chuẩn xác cao, bước đầu xác định chính xác quá trình tiến hoá của loài người, cũng như quá trình tiến hoá của các sinh vật trên trái đất. Ví dụ, từ các dữ liệu phân tích DNA và di truyền miễn dịch, V. M. Sarich và A. C. Wilson (năm 1967) xác định loài người đã tách khỏi tổ tiên chung với tinh tinh từ cách đây hơn 5 triệu năm.

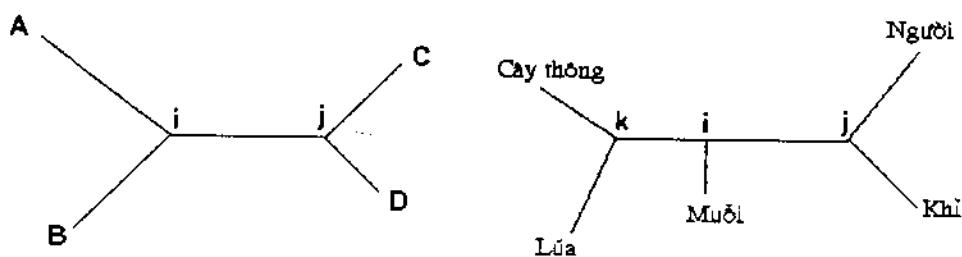
Cây phân loại cho thấy chiều hướng tiến hoá của các loài sinh vật, quan hệ trong nội bộ loài hoặc giữa các loài sinh vật trong hệ thống phân loại. Cây phân loại góp phần giúp các nhà khoa học xây dựng các phép lai có hiệu quả kinh tế cao, chẩn đoán ưu thế lai sớm hoặc tìm kiếm nguồn gốc phát sinh của các loài sinh vật...

## II. Các dạng cây phân loại

Trên cơ sở phân tích dữ liệu di truyền, để đánh giá mức độ giống nhau hoặc khác nhau giữa các quần thể, các loài sinh vật khác nhau người ta sử dụng nhiều phương pháp khác nhau để xây dựng các cây phân loại.

### 1. Cây phân loại không gốc

Cây phân loại được chia làm hai nhóm chủ yếu là cây không gốc và cây có gốc. Cây phân loại không gốc gồm các cành xuất phát từ một cành chung, từ cành chung xuất hiện các nhánh cấp 1, từ nhánh cấp 1 xuất phát các nhánh cấp 2, 3..., các lá xuất phát từ các nhánh tạo nên nhiều dạng hình khác nhau. Trong cành chung, khoảng cách giữa các cành cấp 1 thường không được xác định rõ ràng, hay không xác định chính xác nguồn gốc xuất phát của các nhánh. Ví dụ, trong hình 3.2 các khoảng cách giữa i, j và k không xác định được chính xác.



*Hình 3. 2. Hình dạng cây phân loại không gốc*

## 2. Cây phân loại có gốc

Cây phân loại có gốc gồm các cành xuất phát từ một gốc chính, các nhánh cấp 1, 2,... xuất phát từ các cành, các lá xuất phát từ các nhánh. Sự khác nhau của các đối tượng so sánh về cấu trúc dữ liệu, tạo nên các cây phân loại hình dạng khác nhau. Cùng một số đối tượng so sánh, có thể tạo nên nhiều dạng cây phân loại khác nhau, phụ thuộc vào cấu trúc dữ liệu thu được và quan hệ giữa các đối tượng.

Bảng 3. 1. Số dạng cây và số lượng cành trên một cây phân loại

Dữ liệu		Cây không gốc		Cây có gốc	
Số đối tượng	Khoảng cách	Số dạng cây có thể có	Số cành một cây	Số dạng cây có thể có	Số cành một cây
3	3	1	3	3	4
4	6	3	5	15	6
5	10	15	7	105	8
n	$\frac{n(n-1)}{2}$	$\frac{(2n-5)!}{2^{n-3}(n-3)!}$	$2n - 3$	$\frac{(2n-3)!}{2^{n-2}(n-2)!}$	$2n - 2$

Trong nửa đầu thế kỉ 20, bằng các dữ liệu phân tích tần số allele của các enzym, izoenzym, các tính trạng... người ta đã xây dựng cây phân loại của nhiều loài sinh vật khác nhau, xác định một số quan hệ giữa các loài sinh vật trong tự nhiên.

Hiện nay, từ dữ liệu nghiên cứu trình tự nucleotid trong gen, trình tự acid amin trong phân tử protein con người đã và đang hoàn thiện cây phân loại của nhiều loài virus, vi khuẩn, thực vật, động vật... với độ chuẩn xác cao, bước đầu xác định chính xác quá trình tiến hoá của loài người, cũng như quá trình tiến hoá của các sinh vật trên trái đất. Ví dụ, từ các dữ liệu phân tích DNA và di truyền miễn dịch, V. M. Sarich và A. C. Wilson (năm 1967) xác định loài người đã tách khỏi tổ tiên chung với tinh tinh từ cách đây hơn 5 triệu năm.

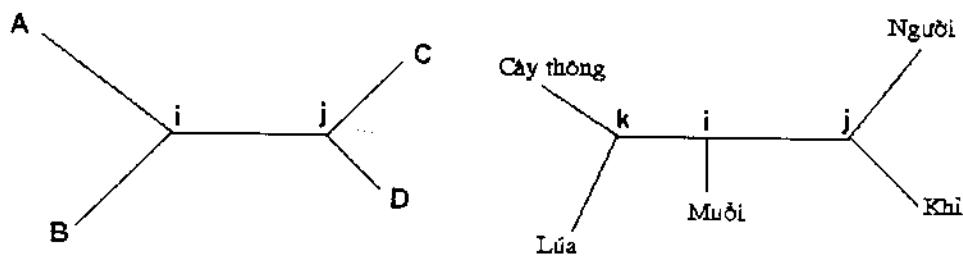
Cây phân loại cho thấy chiều hướng tiến hoá của các loài sinh vật, quan hệ trong nội bộ loài hoặc giữa các loài sinh vật trong hệ thống phân loại. Cây phân loại góp phần giúp các nhà khoa học xây dựng các phép lai có hiệu quả kinh tế cao, chẩn đoán ưu thế lai sớm hoặc tìm kiếm nguồn gốc phát sinh của các loài sinh vật...

## II. Các dạng cây phân loại

Trên cơ sở phân tích dữ liệu di truyền, để đánh giá mức độ giống nhau hoặc khác nhau giữa các quần thể, các loài sinh vật khác nhau người ta sử dụng nhiều phương pháp khác nhau để xây dựng các cây phân loại.

### I. Cây phân loại không gốc

Cây phân loại được chia làm hai nhóm chủ yếu là cây không gốc và cây có gốc. Cây phân loại không gốc gồm các cành xuất phát từ một cành chung, từ cành chung xuất hiện các nhánh cấp 1, từ nhánh cấp 1 xuất phát các nhánh cấp 2, 3..., các lá xuất phát từ các nhánh tạo nên nhiều dạng hình khác nhau. Trong cành chung, khoảng cách giữa các cành cấp 1 thường không được xác định rõ ràng, hay không xác định chính xác nguồn gốc xuất phát của các nhánh. Ví dụ, trong hình 3.2 các khoảng cách giữa i, j và k không xác định được chính xác.



Hình 3.2. Hình dạng cây phân loại không gốc

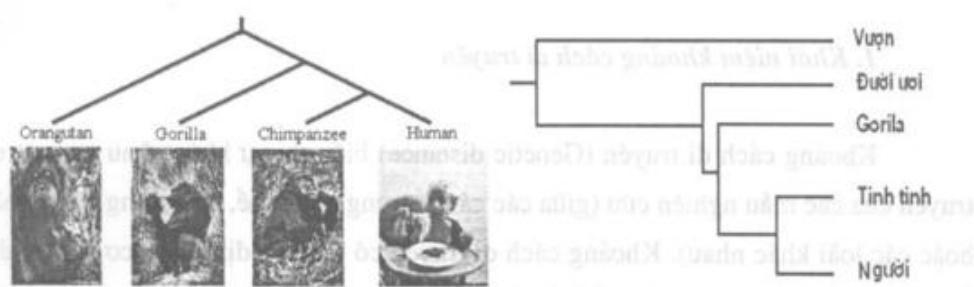
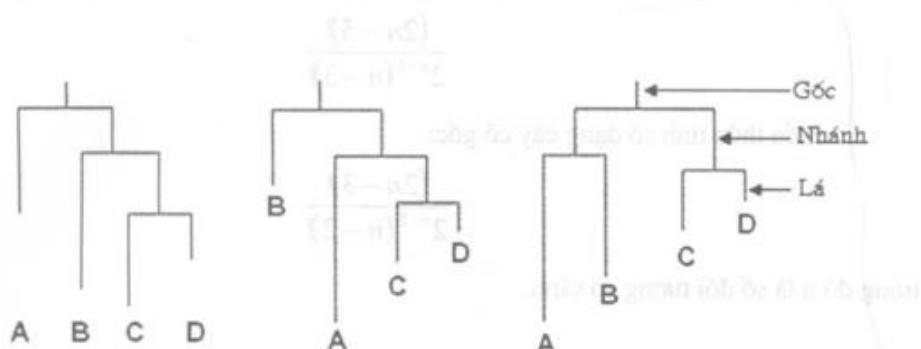
## 2. Cây phân loại có gốc

Cây phân loại có gốc gồm các cành xuất phát từ một gốc chính, các nhánh cấp 1, 2,... xuất phát từ các cành, các lá xuất phát từ các nhánh. Sự khác nhau của các đối tượng so sánh về cấu trúc dữ liệu, tạo nên các cây phân loại hình dạng khác nhau. Cùng một số đối tượng so sánh, có thể tạo nên nhiều dạng cây phân loại khác nhau, phụ thuộc vào cấu trúc dữ liệu thu được và quan hệ giữa các đối tượng.

Bảng 3.1. Số dạng cây và số lượng cành trên một cây phân loại

Dữ liệu		Cây không gốc		Cây có gốc	
Số đối tượng	Khoảng cách	Số dạng cây có thể có	Số cành một cây	Số dạng cây có thể có	Số cành một cây
3	3	1	3	3	4
4	6	3	5	15	6
5	10	15	7	105	8
n	$\frac{n(n-1)}{2}$	$\frac{(2n-5)!}{2^{n-3}(n-3)!}$	$2n - 3$	$\frac{(2n-3)!}{2^{n-2}(n-2)!}$	$2n - 2$

Ví dụ, có 4 trình tự nucleotid kí hiệu A, B, C và D có thể dựng được tối đa có số dạng cây phân loại có gốc là 15 (bảng 3.1.). Hình dáng của mỗi dạng cây khác nhau tuỳ thuộc đặc điểm khác nhau giữa các trình tự A, B, C và D (hình 3.3).



Hình 3.3. Các dạng cây phân loại có gốc

### 3. Số lượng và số dạng cây phân loại

Số dạng cây phân loại có thể dựng được tuỳ thuộc cấu trúc của dữ liệu di truyền, số lượng đối tượng so sánh và dạng cây phân loại cần xây dựng (dạng cây không gốc hoặc cây có gốc). Số đối tượng so sánh càng lớn ( $n$  càng lớn) thì số dạng cây có thể có càng lớn, và ngược lại. Căn cứ vào số lượng đối tượng so sánh, có thể tính được số cành trên mỗi cây và số lượng dạng cây phân loại có thể có.

Ví dụ, có 5 đối tượng so sánh ( $n = 5$ ,  $n$  có thể là một trình tự, một gen, một locus gen, một cá thể hoặc một quần thể... ). Nếu xây dựng cây phân loại không gốc,

tối đa có thể dùng được 15 dạng cây. Nếu dùng cây phân loại có gốc, có thể dùng được 105 dạng cây khác nhau (bảng 3.1).

Số dạng cây không gốc có thể tính theo biểu thức:

$$\frac{(2n - 5)!}{2^{n-3}(n - 3)!}$$

Biểu thức tính số dạng cây có gốc:

$$\frac{(2n - 3)!}{2^{n-2}(n - 2)!}$$

trong đó n là số đối tượng so sánh.

### **III. Khoảng cách di truyền**

#### *1. Khái niệm khoảng cách di truyền*

Khoảng cách di truyền (Genetic distance) biểu thị sự khác nhau về mặt di truyền của các mẫu nghiên cứu (giữa các cá thể trong quần thể, các giống trong loài hoặc các loài khác nhau). Khoảng cách di truyền có thể xác định trên cơ sở các dữ liệu di truyền khác nhau như tần số các allele trong một locus, trình tự sắp xếp các nucleotid trong gen, trình tự các acid amin trong phân tử protein hoặc kích thước số lượng các đoạn DNA tạo được bằng các kỹ thuật phân tử RFLP, tách dòng với BAC, YAC...

Căn cứ vào cấu trúc dữ liệu di truyền thu được từ các nghiên cứu là trình tự nucleotid của gen, trình tự acid amin trong chuỗi polypeptid hoặc tần số allele trong các locus..., người ta đã xây dựng nên các phương pháp tính khoảng cách di truyền khác nhau. Do đó, khi tính toán khoảng cách di truyền trong các nghiên cứu, cần lựa chọn phương pháp xử lý số liệu phù hợp với cấu trúc dữ liệu, để đảm bảo kết quả có độ chính xác cao.

Trong tài liệu này, chúng tôi giới thiệu một số biểu thức chủ yếu của nhiều tác giả khác nhau, được sử dụng tương đối phổ biến và được công bố trên các tạp chí chuyên ngành uy tín. Mỗi biểu thức tính khoảng cách di truyền hoặc phương

pháp dụng cây phân loại giữa các quần thể nghiên cứu, cho kết quả chính xác trong những điều kiện nhất định. Cùng một dữ liệu di truyền, khi sử dụng các phương pháp khác nhau có thể thu được các kết quả có sự khác biệt nhau ít nhiều, dẫn đến hình dạng và cấu trúc cây phân loại có thể khác nhau (Mashurov A.M, Thanh K.H, và cộng sự 1992, 1998). Mỗi kiểu dữ liệu di truyền thu được từ các phương pháp nghiên cứu khác nhau (tần số các allele, kích thước đoạn DNA, trình tự nucleotid...) cần các phương pháp xử lý số liệu khác nhau. Do đó, trong nghiên cứu sinh học để có kết quả chính xác nhất, cần phải lựa chọn biểu thức tính khoảng cách di truyền, phương pháp dựng cây phân loại phù hợp với cấu trúc dữ liệu thu được từ các thực nghiệm.

## ***2. Phương pháp xác định khoảng cách di truyền với dữ liệu trình tự các nucleotid trong gen và bộ gen***

Xác định khoảng cách di truyền với dữ liệu là trình tự của các nucleotid, có vai trò quan trọng trong nghiên cứu tiến hóa phân tử, giúp con người xác định nguồn gốc và sự tiến hóa của sinh vật chính xác.

Các trình tự nucleotid được sử dụng để xác định khoảng cách di truyền có thể là trình tự nucleotid mang mã di truyền (coding sequence), các trình tự nucleotid không mang mã di truyền (non coding sequence), hoặc các đoạn trình tự lặp lại trong cấu trúc bộ gen... Đến nay, có nhiều phương pháp xác định khoảng cách di truyền từ dữ liệu trình tự nucleotid, trong đó phổ biến nhất là các phương pháp: Jukes và Cantor (1969), Kimura (1980), Tajima và Nei (1984); Tamura và Nei (1993), F84 (Felsenstein 1993), và Lake (1994)...

Các phương pháp Tamura và Nei (1993), F84 (Felsenstein 1993) được sử dụng xây dựng nhiều phần mềm tính khoảng cách di truyền trên máy tính (Xunhua Xia, 2003).

### **2. 1. Xác định khoảng cách di truyền theo Jukes và Cantor - $D_{JC}$ (1969)**

$$D_{JC} = -\frac{3}{4} \cdot \ln \left( 1 - \frac{4p}{3} \right)$$

trong đó p là sự khác biệt giữa 2 trình tự.

### **2. 2. Xác định khoảng cách di truyền theo Tajima và Nei – $D_{TN}$ (1984)**

$$b = \frac{\left(1 - \sum_{i=1}^4 q_i^2\right) + \left(\frac{p^2}{h}\right)}{2}$$

$$h = \sum_{i=1}^3 \sum_{j=i+1}^4 \frac{x_{ij}^2}{2q_i q_j}$$

$$D_{TN} = \frac{b^2 \cdot p(1-p)}{L(b-p)^2}$$

trong đó,  $x_{ij}$  là tỉ lệ các cặp bazơ thứ i và j giữa 2 trình tự so sánh,  $q_i$  là tần số của nucleotid thứ i ( $i = 1, 2, 3, 4$  tương ứng với A, G, T, C), L là chiều dài chuỗi nucleotid.

### **2. 3. Xác định khoảng cách di truyền theo Kimura (1980)**

$$D_{K80} = \frac{\ln(a)}{2} + \frac{\ln(b)}{4}$$

Đặt P là tỉ lệ khác nhau giữa 2 trình tự, Q là sự khác nhau theo hàng ngang giữa 2 trình tự. P và Q nằm trong khoảng 0,25 đến 0,5. Biểu thức Kimura biểu diễn dưới dạng:

$$D_{K80} = \frac{a^2 P + c^2 Q - (aP + cQ)^2}{L}$$

$$a = \frac{1}{1 - 2P - Q}$$

$$b = \frac{1}{1 - 2Q}$$

$$c = \frac{a + b}{2}$$

#### 2. 4. Xác định khoảng cách di truyền theo Tamura và Nei - $D_{TN}$ (1993)

$$D_{TN93} = 4\pi_T\pi_C K_1 + 4\pi_A\pi_G K_2 + 4\pi_Y\pi_R \beta$$

$$K_1 = \frac{-\ln\left(1 - \frac{\pi_Y P_1}{2\pi_T\pi_C} - \frac{Q}{2\pi_Y}\right) - \pi_R \ln\left(1 - \frac{q}{2\pi_Y\pi_R}\right)}{2\pi_Y}$$

$$K_2 = \frac{-\ln\left(1 - \frac{\pi_R P_2}{2\pi_A\pi_G} - \frac{Q}{2\pi_R}\right) - \pi_Y \ln\left(1 - \frac{q}{2\pi_Y\pi_R}\right)}{2\pi_R}$$

$$\beta = \frac{-\ln\left(1 - \frac{Q}{2\pi_Y\pi_R}\right)}{2}$$

#### 2. 5. Xác định khoảng cách di truyền F 84 (PHYLIP 1984)

F 84 là biểu thức được sử dụng trong phần mềm tính khoảng cách di truyền của các trình tự nucleotid trên DNA được xây dựng năm 1984 (PHYLIP, 1984).

$$D_{F84} = 4\beta \left( \pi_T\pi_C * \left( 1 + \frac{k}{\pi_Y} \right) + \pi_A\pi_G * \left( 1 + \frac{k}{\pi_R} \right) + \pi_Y * \pi_R \right)$$

$$\beta = \frac{-\ln\left(1 - \frac{Q}{2\pi_Y\pi_R}\right)}{2}$$

$$k = \frac{\alpha}{\beta} - 1$$

$$\alpha = \frac{-\ln \left( -\frac{2(\pi_T \pi_C + \pi_A \pi_G) + 2 \left( \frac{\pi_T \pi_C \pi_R}{\pi_Y} + \frac{\pi_A \pi_G \pi_Y}{\pi_R} \right) \left( 1 - \frac{Q}{2\pi_Y \pi_c} \right) - P}{2\pi_T \pi_C / \pi_Y + 2\pi_A \pi_G / \pi_R} \right)}{2}$$

### *3. Phương pháp xác định khoảng cách di truyền trên cơ sở các dấu chuẩn vi vệ tinh trong gen và bộ gen*

Trong gen hoặc bộ gen của sinh vật bậc cao có hiện tượng lặp lại nhiều lần một trình tự nucleotid nhất định, kí hiệu là VNTR (Variable Number of Tandem Repeat) hoặc SSR (simple sequence repeats). Các trình tự lặp (VNTR) trong các gen tạo nên các dãy da alell, trình tự VNTR trong bộ gen, tạo nên sự đa hình di truyền của các loài sinh vật. VNTR gồm hai dạng chính là minisatellit và microsatellit. Minisatellit là những đoạn lặp trình tự khoảng 15 - 500 bp, còn microsatellit là những đoạn lặp một trình tự gồm 2 - 6 bp. Bộ gen sinh vật bậc cao có thể lặp lại một số lần đến hàng trăm lần lặp lại một trình tự nào đó. Ví dụ, nhiễm sắc thể X của người có 5 gen có các microsatellit kích thước 300 - 500 kb (Edward và cộng sự, 1991).

Dấu chuẩn vi vệ tinh (microsatellit) được sử dụng như một dấu chuẩn (marker) trong phân tích di truyền. Nhiều tác giả sử dụng microsatellit để xác định nguồn gốc và sự khác biệt giữa các cá thể, hoặc giữa các loài trong nghiên cứu tiến hóa của sinh vật. Trên cơ sở dữ liệu phân tích microsatellit, nhiều nhà khoa học đã xây dựng các công thức khác nhau để xác định khoảng cách di truyền giữa các quần thể, các loài sinh vật (Cavalli - Sforza và Edwards, 1967, 1971; Raynolds và cộng sự, 1983; Nei, 1978, 1984; Gottelli và cộng sự, 1994; Lade và cộng sự, 1996....). Ở Việt Nam, nhiều tác giả đã sử dụng các chỉ thị microsatellit hoặc VNTR để đánh giá đa dạng sinh học hoặc phân loại phân tử. Ví dụ, đánh giá các quần thể đậu tương cúc vàng (Trần Thị Phương Liên và cộng sự, 2004), đánh giá đa hình di truyền các quần thể tôm sú (Nguyễn Thị Thảo và cộng sự, 2004)...

Trong tài liệu này, chúng tôi giới thiệu một số phương pháp tính khoảng cách di truyền giữa các quần thể có các dãy đa alell do microsatellit tạo nên, tạo cơ sở cho các lựa chọn khi nghiên cứu đa hình di truyền các đoạn DNA lặp lại trong gen và bộ gen của sinh vật bậc cao.

3. 1. Phương pháp tính khoảng cách di truyền giữa các quần thể có các alell microsatellit theo Cavalli - Sforza và Bodmer (1971)

$$D_{kf} = -\ln (kf)$$

$$kf = \sum_{i=1}^n P[A(i)] \cdot P[B(i)]$$

trong đó:  $P[A(i)]$  là tần số tương đối của alell thứ i trong quần thể A;

$P[B(i)]$  là tần số tương đối của alell thứ i trong quần thể B

3. 2. Phương pháp tính khoảng cách di truyền giữa các quần thể có các alell microsatellit theo Nei (1978)

$$Ds = -\ln (id)$$

$$id = \frac{\sum_{i=1}^n P[A(i)] \times P[B(i)]}{\sqrt{\left( \sum_{i=1}^n P[A(i)] \right)^2 \times \left( \sum_{i=1}^n P[B(i)] \right)^2}}$$

3. 3. Phương pháp tính khoảng cách di truyền giữa các quần thể có các alell microsatellit theo Raynolds và công sự (1983)

$$D_{Fst} = -\ln (1 - Theta)$$

$$\text{Theta} = \frac{a}{a + b}$$

trong đó, a là số lượng các allele trong dãy da allele tạo thành bởi các microsatellit của quần thể A, b là số lượng các allele tương ứng của quần thể B.

#### *4. Phương pháp xác định khoảng cách di truyền với dữ liệu trình tự acid amin trong chuỗi polypeptid*

Phương pháp xác định khoảng cách di truyền với dữ liệu trình tự các acid amin trong chuỗi polypeptid, được sử dụng phổ biến là phương pháp Dayhoff (1978) và phương pháp Kimura (1983). Các phương pháp này có biểu thức tính khoảng cách di truyền giữa 2 chuỗi polypeptid:

$$D = -\ln(1 - p - 0,2p^2)$$

trong đó, p số acid amin khác nhau giữa 2 chuỗi polypeptid. Do các chuỗi peptid không chỉ khác nhau về số lượng và số loại acid amin, còn khác nhau ở tính chất vật lí và hoá học của các loại acid amin trong cấu trúc chuỗi peptid. Khi tính khoảng cách di truyền giữa các acid amin, cần chú ý sự khác nhau về cấu trúc hoá học của các acid amin trong chuỗi polypeptid.

#### *5. Phương pháp xác định khoảng cách di truyền với dữ liệu tần số allele của các locus gen*

Có 2 quần thể X và Y. Trong đó:

$X_{ij}, Y_{ij}$ : tần số của allele thứ i trong locus thứ j của các quần thể so sánh;

$k_i$ : tổng số allele trong j locus;

n : số locus nghiên cứu;

m : số cặp allele nghiên cứu tổng số các đặc điểm trong n locus );

N : số cá thể trong các quần thể so sánh.

Một số phương pháp tính khoảng cách di truyền giữa các locus gen, được sử dụng trong xây dựng các phần mềm máy tính như phương pháp Nei (1972); Nei (1983); Cavalli-Sforza (1967); Reynolds và cộng sự (1983)...

5. 1. Tính khoảng cách di truyền theo Smith (1957); Sokal (1981)

$$d_s = \sqrt{\frac{1}{\sum_{i=1}^n k_i} \sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} - Y_{ij})^2}$$

5. 2. Tính khoảng cách di truyền theo Cavalli - Sforza và Edward (1967)

$$d_{cs} = \frac{2\sqrt{2}}{\pi} \cdot \sqrt{\sum_{i=1}^n \left( 1 - \sum_{j=1}^{k_i} \sqrt{X_{ij} \cdot Y_{ij}} \right)}$$

5. 3. Tính khoảng cách di truyền theo Rogers (1972)

$$d_R = \frac{1}{n} \sum_{i=1}^n \sqrt{\frac{1}{2} \sum_{j=1}^{k_i} (X_{ij} - Y_{ij})^2}$$

5. 4. Tính khoảng cách di truyền theo Kidd (1974)

$$d_{Ki} = 4 \cdot \frac{1}{\sum_{i=1}^n k_i - 1} \cdot \sum_{i=1}^n \left( 1 - \sum_{j=1}^{k_i} \sqrt{X_{ij} \cdot Y_{ij}} \right)$$

5. 5. Tính khoảng cách di truyền theo Nei (1972)

$$d_N = -\ln \sqrt{\left( \sum_{i=1}^n \sum_j X_{ij}^2 \right) \left( \sum_{i=1}^n \sum_j Y_{ij}^2 \right)} \sum_{i=1}^n \sum_{j=1}^{k_i} (X_{ij} \cdot Y_{ij})$$

**5. 6. Tính khoảng cách di truyền theo Nei (1983)**

$$d_{N83} = \frac{\sum_{i=1}^n \left( 1 - \sum_{j=1}^{ki} \sqrt{X_{ij} \cdot Y_{ij}} \right)}{m}$$

**5. 7. Tính khoảng cách di truyền theo Hedrick (1975)**

$$d_H = 1 - \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^{ki} (X_{ij} - Y_{ij})^2}{\frac{1}{2} \left( \sum_{j=1}^{ki} X_{ij}^2 + \sum_{j=1}^{ki} Y_{ij}^2 \right)}$$

**5. 8. Tính khoảng cách di truyền theo Reynolds, và công sự (1983)**

$$d_{RWC} = \frac{\sum_{i=1}^n \sum_{j=1}^{ki} (X_{ij} - Y_{ij})^2}{2 \sum_{i=1}^n \left( 1 - \sum_{j=1}^{ki} X_{ij} \cdot Y_{ij} \right)}$$

**5. 9. Tính khoảng cách di truyền theo Jivotovskovo và Mashurov (1974)**

$$d_{ZM} = 1 - \prod_{i=1}^n \sum_{j=1}^{ki} \sqrt{X_{ij} \cdot Y_{ij}}$$

**5. 10. Tính khoảng cách di truyền theo Eisner và Mesejakov (1978)**

$$d_{EM} = \frac{\sum_{i=1}^n \sum_{j=1}^{ki} (di.100)}{n.N.V}$$

5. 11. Phương pháp tính khoảng cách di truyền theo Mesejiaakov (1983)

$$d_M = 1 - \frac{\left( \sum_{i=1}^n \sum_{j=1}^{ki} X_{ij} + \sum_{i=1}^n \sum_{j=1}^{ki} Y_{ij} \right) - \sum_{i=1}^n \sum_{j=1}^{ki} |X_{ij} - Y_{ij}|}{\sum_{i=1}^n \sum_{j=1}^{ki} X_{ij} + \sum_{i=1}^n \sum_{j=1}^{ki} Y_{ij}}$$

5. 12. Phương pháp tính khoảng cách di truyền theo phương pháp UPGMA (Unweighted Pair Group Method Using Arithmetic Averages, Sokal và Michener, 1958)

Khoảng cách di truyền  $d_{ij}$  giữa hai quần thể X và Y (X, Y có thể là hai quần thể, hoặc 2 nhóm khác nhau... ).

$$d_y = \frac{1}{|C_i| \cdot |C_j|} \sum_{p \in C_i, q \in C_j} d_{pq}$$

trong đó:  $|C_i|$  và  $|C_j|$  là giá trị tuyệt đối các trị số của cluster thứ i và j. Giữa 2 nhóm (cluster) chọn các trị số i và j sao cho khoảng cách di truyền  $d_{ij}$  là nhỏ nhất. Chọn một nhóm k sao cho  $C_k = C_i \cup C_j$ , tìm khoảng cách  $d_{kl}$  với mọi l (với  $C_l$  là một nhóm bất kỳ).

5. 13. Phương pháp tính khoảng cách di truyền theo phương pháp NJ - Neighbor joining ( Saitou và Nei )

$$S_{mn} = \frac{\sum d_{im} + d_{in}}{2(N-2)} + \frac{d_{mn}}{2} + \frac{\sum d_{ij}}{(N-2)}$$

$$d_{im} + d_{ik} = d_{ik} \quad d_{jn} + d_{nk} = d_{jk} \quad d_{in} + d_{jn} = d_{ij}$$

$$d_{in} = (d_{ij} + r_i - r_j)/2 \quad d_{jn} = (d_{ij} + r_j - r_i)/2$$

#### **IV. Phương pháp dựng cây phân loại với dữ liệu khoảng cách di truyền**

##### *1. Nguyên tắc dựng cây phân loại*

Cây phân loại có thể được xây dựng bằng nhiều phương pháp khác nhau, trong đó thông dụng nhất là 2 nhóm phương pháp: phương pháp khoảng cách di truyền (distance method) và phương pháp trên cơ sở các kí tự (character - based method).

Dụng cây phân loại dựa trên dữ liệu khoảng cách di truyền của các đối tượng nghiên cứu, gọi là phương pháp khoảng cách di truyền (distance methods). Phương pháp khoảng cách di truyền là phương pháp xuất hiện sớm nhất, được sử dụng rộng rãi với nhiều loại đối tượng nghiên cứu khác nhau, cho phép xây dựng cây phân loại với tập hợp dữ liệu rất lớn. Các dữ liệu di truyền về tần số các allele của các protein, enzym, hoặc các trình tự nucleotid, kích thước các đoạn DNA, RNA hoặc trình tự acid amin trong chuỗi peptid... đều có thể được sử dụng để dựng cây phân loại.

Cây phân loại xây dựng theo phương pháp khoảng cách di truyền, gồm nhiều nhóm phương pháp khác nhau, xây dựng trên các dữ liệu di truyền khác nhau. Hiện nay, có nhiều phương pháp được xây dựng bởi các tác giả khác nhau: Phương pháp UPGMA (Unweighted Pair Group Method use Arithmetic); Phương pháp NJ (Neighbour Joining - cải tiến từ phương pháp UPGMA của Saitou và Nei, 1987); Phương pháp tiến hoá nhỏ (ME - minimum evolution); Phương pháp (FM - Fitch-Margoliash, 1967)... Trong đó, hai phương pháp được sử dụng phổ biến nhất là UPGMA và NJ. Phương pháp UPGMA có ưu điểm dễ sử dụng (Helsenbeck, 1995).

còn phương pháp NJ được đánh giá có độ chuẩn xác cao hơn các phương pháp UPGMA và ME (Rzhetsky và Nei, 1992; Li, 1997). Tuy nhiên, mức độ chuẩn xác của phương pháp, sự lựa chọn phương pháp tuỳ thuộc cấu trúc dữ liệu và thói quen của người sử dụng.

Phương pháp dựng cây phân loại trên cơ sở các kí tự (character-based method) gồm các phương pháp thông dụng như cực đại parsimony (MP - maximum parsimony) và phương pháp tiếp cận cực đại khả năng có thể xảy ra (ML - maximum likelihood). Các phương pháp này dựa vào sự thay đổi của các kí tự trong chuỗi nucleotid hoặc sự thay đổi vị trí các loại acid amin trong chuỗi polypeptid, để xếp các đối tượng có quan hệ gần gũi với nhau nhất thành các nhóm, từ quan hệ giữa các nhóm để xây dựng cây phân loại.

Hiện nay, có nhiều phần mềm giúp con người phân tích dữ liệu di truyền và dựng cây phân loại trên máy tính. Mỗi phần mềm được lập trình trên cơ sở lý thuyết của một hoặc một số phương pháp phân tích di truyền nhất định. Ví dụ, phần mềm METREE (Rzhetsky và Nei, 1994) được lập trình dựa trên phương pháp ME (minimum evolution); phần mềm PHYLIP được lập trình trên cơ sở phương pháp FM (Fitch-Margoliash) và cơ sở dữ liệu protein...

Phần mềm PHYLIP (PHYLogeny Inference Package) do J. Felsenstein và cộng sự lập trình (1993), là một trong những phần mềm được ứng dụng rộng rãi. Ngoài ra, có rất nhiều phần mềm khác, giúp con người xử lí số liệu di truyền và dựng cây phân loại trên máy tính nhanh, chính xác như CLUSTAL W, PileUp, ALIGN in ProPack, SAM (Hughey et al., 1996) và MACAW (Lawrence et al., 1993), NTSYSpc 2.0...

Tuy nhiên, các phần mềm xử lí dữ liệu di truyền và dựng cây phân loại trên máy tính chỉ có ý nghĩa là công cụ, hỗ trợ con người nhanh chóng đạt mục đích, phân tích số liệu nhanh và chính xác. Để lựa chọn một phần mềm phù hợp và sử dụng các phần mềm có hiệu quả, người sử dụng cần nắm vững nguyên lí cơ bản

của phương pháp đã lập trình nên mỗi phần mềm. Do đó, cần căn cứ vào loại dữ liệu di truyền và độ lớn của tập hợp dữ liệu để lựa chọn phần mềm phù hợp.

**2. Các bước xây dựng cây phân loại với dữ liệu khoảng cách di truyền của các trình tự DNA, RNA, protein...**

**Bước 1:**

Từ nghiên cứu bộ gen (hoặc gen) của các đối tượng, thu nhận được dữ liệu di truyền là các trình tự nucleotid. Chú ý, chọn cùng loại đối tượng nghiên cứu. Ví dụ, cùng là đoạn gen mã hoá enzym lipase, cùng là đoạn gen cytochrom C trong ti thể..., chọn các trình tự kích thước tương đương nhau (số lượng nucleotid, ribonucleotid, acid amin ).

**Bước 2:**

Sắp xếp lại các trình tự dữ liệu, làm cho điểm xuất phát đầu tiên giống nhau. Ví dụ, bắt đầu từ nucleotid đầu tiên được phiên mã, từ bộ ba mở đầu dịch mã, hoặc bắt đầu từ vị trí cắt của cùng một loại enzym giới hạn...

**Bước 3:**

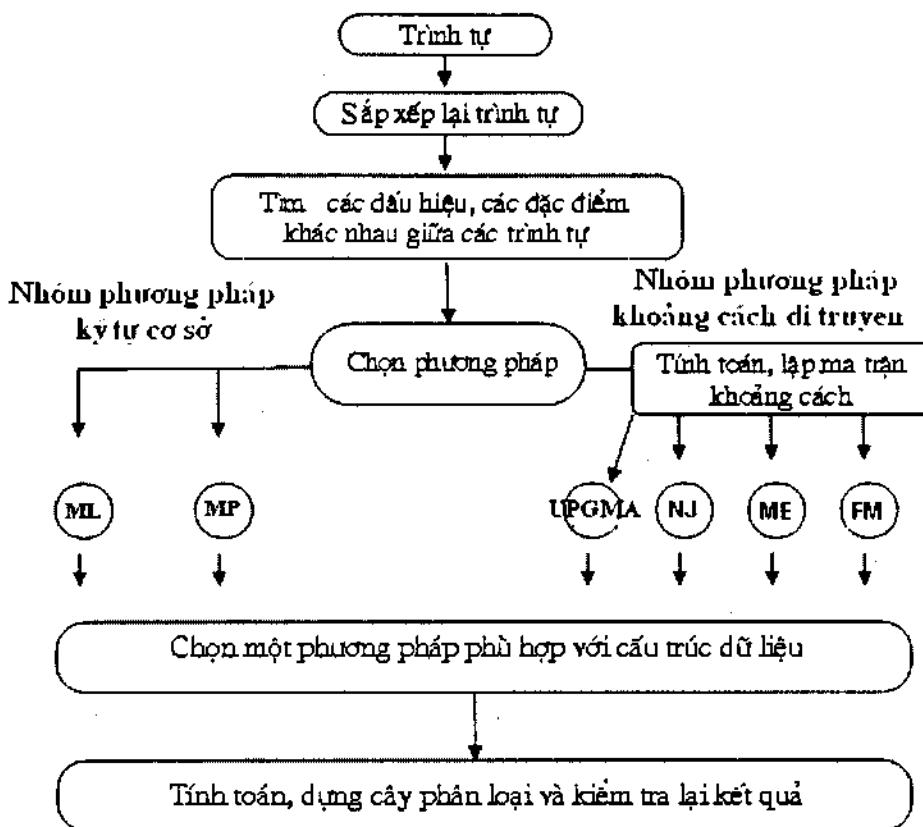
So sánh giữa các trình tự để tìm kiếm các đặc điểm, các tín hiệu khác nhau giữa các trình tự nghiên cứu (sự khác nhau về khoảng cách di truyền, tần số allele hoặc khác nhau giữa các kí tự...)

**Bước 4:**

Lựa chọn phương pháp dựng cây phân loại thuộc nhóm phương pháp khoảng cách di truyền hoặc nhóm phương pháp kí tự cơ sở. Ví dụ, chọn phương pháp UPGMA, NJ, ME...

**Bước 5:**

Tính toán kết quả, từ các số liệu thu được xây dựng ma trận khoảng cách, sau đó dựng cây phân loại và kiểm tra độ chuẩn xác của các dữ liệu.



Hình 3. 4. Sơ đồ các bước dựng cây phân loại

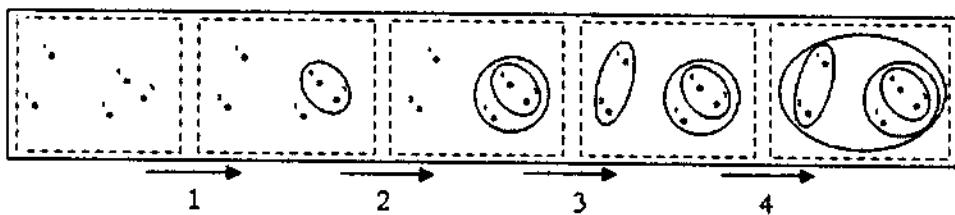
### 3. Phương pháp lập ma trận khoảng cách di truyền và dựng cây phân loại

Phương pháp dựng cây phân loại dựa vào khoảng cách di truyền (còn gọi là phương pháp khoảng cách di truyền - distance method) là phương pháp dựng cây phân loại phổ biến và có hiệu quả cao, cho phép phân tích một khối lượng dữ liệu lớn, độ chính xác cao.

Từ các dữ liệu thu được sự khác nhau về khoảng cách di truyền giữa các quần thể, xây dựng được ma trận khoảng cách di truyền (distance matrix). Ma trận

(mẫu nghiên cứu có thể là tần số allele, các trình tự nucleotid, các đoạn DNA... ). So sánh sự khác biệt giá trị giữa các đối tượng trong ma trận thứ nhất, chọn giá trị khoảng cách ( $d_{ij}$ ) giữa 2 quần thể nhỏ nhất làm cụm xuất phát (cluster 1). Sau đó, tính các giá trị khác biệt của cụm xuất phát với các mẫu còn lại được ma trận khoảng cách di truyền thứ 2, so sánh để chọn cụm thứ 2 (cluster 2) có giá trị nhỏ nhất trong ma trận khoảng cách thứ 2. Tiếp tục tính toán đến cụm cuối cùng, cụm thứ m (cluster m) bằng cách lần lượt chọn các khoảng cách nhỏ nhất.

Ví dụ, hình 3.5 biểu hiện vị trí phân bố không gian theo khoảng cách di truyền của 5 quần thể nghiên cứu (ký hiệu 1, 2, 3, 4 và 5). Dựa vào khoảng cách di truyền, giữa quần thể số 4 và quần thể số 5 có khoảng cách di truyền nhỏ nhất, chọn cụm thứ nhất là (4 - 5). Giữa cụm thứ nhất (4-5) với quần thể số 3 có khoảng cách nhỏ nhất, ta có cụm thứ hai{(4 - 5)3}. Hai quần thể 1 và số 2 có khoảng cách nhỏ hơn giữa các quần thể khác, chọn cụm thứ ba (1- 2) là cụm giữa 2 quần thể số 1 và số 2. Cụm cuối cùng là cụm giữa cụm thứ 2 và cụm thứ 3: {(1 - 2) - (4 - 5)3}. Tập hợp các kết quả phân tích trên, có thể dựng được cây phân loại giữa các quần thể.



**Hình 3.5. Sơ đồ phân bố không gian của các quần thể nghiên cứu theo khoảng cách di truyền**

### 3. 1. Ví dụ 1:

Có 4 quần thể cùng loài kí hiệu A, B, C, và D ( $n = 4$ ), mỗi quần thể gồm 100 cá thể được lựa chọn theo sự khác nhau về màu sắc hoa giữa các quần thể: Quần thể A và quần thể B có 6 cá thể khác nhau, A với C = 5, B với D = 12... biểu

hiện ở bảng dưới đây. Xác lập mối quan hệ giữa các quân thể A, B, C, và D bằng cây phân loại?

	A	B	C	D
A	-			
B	6	-		
C	5	7	-	
D	10	12	7	-

Ký hiệu  $d$  là khoảng cách di truyền giữa các quân thể. Trong ma trận sự sụt khác nhau giữa số lượng cá thể của các quân thể, cho thấy quân thể A và C có sự khác biệt giữa nhỏ nhất (khoảng cách di truyền nhỏ nhất)  $d(AC) = 5$ . Do đó, chọn cụm thứ nhất (cluster 1) là (AC). Tính khoảng cách di truyền giữa cụm thứ nhất: (AC) với các quân thể còn lại theo nguyên tắc:

$$d \{(AC), B\} = 1/2 d(AB) + 1/2 d(CB)$$

$$d \{(AC), D\} = 1/2 d(AD) + 1/2 d(CD)$$

Từ số liệu tính toán thu được, ta có ma trận khoảng cách di truyền thứ hai:

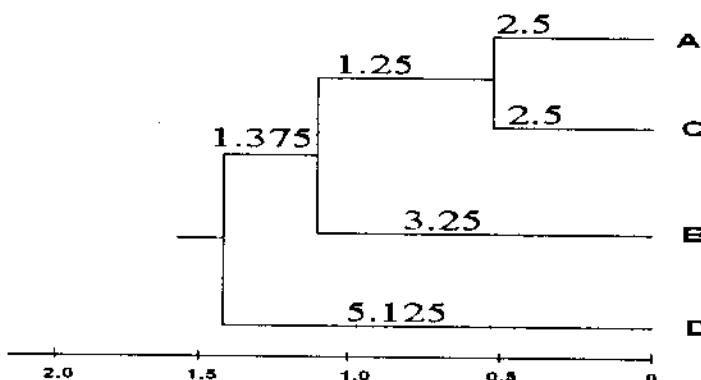
	AC	B	D
AC	-		
B	6,5	-	
D	8,5	12	-

Trong ma trận khoảng cách thứ hai, khoảng cách di truyền giữa quân thể B với cụm thứ nhất là nhỏ nhất  $d \{(AC)\} \text{ với } B\} = 6,5$ . Do đó, chọn được cụm thứ 2 giữa quân thể B với cụm (AC) là:  $\{(AC), B\}$ .

Tính khoảng cách di truyền giữa cụm thứ hai:  $\{(AC), B\}$  với quân thể D theo nguyên tắc:  $d \{(AC), B\} D\} = 1/2 d\{(AC), D\} + 1/2 d \{(B), D\}$ . Thu được ma trận khoảng cách thứ ba:

	(AC)B	D
(AC)B	-	
D	10,25	-

Từ các số liệu thu được, có thể dựng được cây phân loại biểu diễn quan hệ và khoảng cách di truyền giữa các quần thể A, B, C và D. Cây phân loại (hình 3.6) cho thấy: hai quần thể A và C có quan hệ gần nhau nhất, quần thể D có sự khác biệt nhiều nhất so với quần thể A.



*Hình 3. 6. Cây phân loại biểu thị quan hệ giữa 4 quần thể A, B, C và D*

### 3. 2. Ví dụ 2:

Giải trình tự gen mã hoá phân tử 16 S rRNA (RNA riboxom 16S) của 4 chủng vi khuẩn kí hiệu a, b, c và d, thu được 4 trình tự đặc trưng ở mạch làm khuôn cho sự phiên mã (mạch antisense) của gen là:

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Chủng a: A G G C G G A C A A A C C T

Chủng b: A G G A C G C C A A G C C T

Chủng c: A G G A A A T C T A A C C T

Chủng d: A G G A A A G C T A G C C T

Xác định mối quan hệ chủng loại và dựng cây phân loại giữa 4 chủng vi khuẩn nói trên?

Bước 1: Lập ma trận khoảng cách di truyền thứ nhất

So sánh sự khác nhau về trình tự gen mã hoá phân tử 16 S rRNA của 4 chủng vi khuẩn thu được ma trận khoảng cách thứ nhất:

	a	b	c	d
a	-			
b	4	-		
c	5	5	-	
d	6	4	2	-

Bước 2: Lập ma trận khoảng cách di truyền thứ hai

Ma trận khoảng cách di truyền thứ nhất cho thấy khoảng cách di truyền giữa hai chủng c và d nhỏ nhất với  $d(cd) = 2$ , chọn  $(cd)$  làm cụm thứ nhất. Tính khoảng cách di truyền giữa cụm thứ nhất với các chủng vi khuẩn còn lại theo biểu thức:

$$d\{(cd), a\} = 1/2 d(ac) + 1/2 d(ad) = 1/2 (5 + 6) = 5,5$$

$$d\{(cd), b\} = 1/2 d(bc) + 1/2 d(bd) = 1/2 (5 + 4) = 4,5$$

Ta có ma trận khoảng cách di truyền thứ hai:

	(cd)	a	b
(cd)	-		
a	5,5	-	
b	4,5	4	-

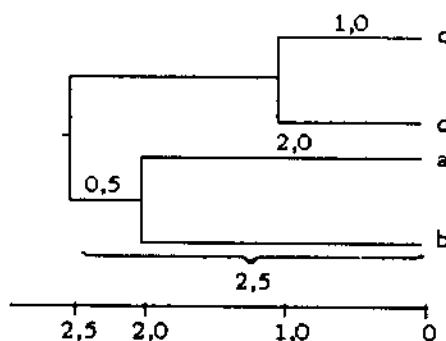
Bước 3: Lập ma trận khoảng cách di truyền thứ ba

Số liệu thu được từ ma trận khoảng cách thứ hai cho thấy khoảng cách di truyền giữa chủng a với chủng b có giá trị nhỏ nhất  $\{d(ab)\} = 4$  tạo thành cụm thứ hai. Khoảng cách di truyền giữa 2 cụm (cd) và (ab) được tính theo biểu thức:

$$d\{(ab), (cd)\} = 1/4 d(ac) + 1/4 d(ad) = 1/4 (5,5 + 4,5) = 2,5$$

Bước 4: Từ số liệu các khoảng cách di truyền đã tính, dựng được cây phân loại biểu thị quan hệ giữa 4 chủng vi khuẩn a, b, c và d:

Cây phân loại (hình 3.7) cho thấy các chủng vi khuẩn nghiên cứu tạo thành 2 nhóm riêng biệt. Hai chủng vi khuẩn c và d có quan hệ gần nhau hơn so với quan hệ di truyền giữa 2 chủng c và d.



*Hình 3.7. Cây phân loại của 4 chủng vi khuẩn a, b, c và d*

Cây phân loại cho thấy 2 chủng c và d xuất phát từ cùng một nguồn gốc, còn 2 chủng a và b xuất phát từ một nguồn gốc khác, có sự khác biệt về mặt di truyền lớn hơn.

### 3. 3. Ví dụ 3:

Năm 1982, Brown và cộng sự nghiên cứu trình tự gen mã hoá RNA vận chuyển histidin (tRNA-His) trong ti thể của 5 đại diện họ linh trưởng: 1- Người (human- hu); 2- Tinh tinh (chimpanzee - ch); 3- Gorila (go); 4- Đười ươi (Orangutan-or); 5- Vượn (Gibbon- gi) thu được các trình tự nucleotid sau:

1-hu: GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA  
CAACAGAGGC TTACGACCCC TTATTTACC

2-ch: GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA  
CAACAGAGGC TCACGACCCC TTATTTACC

3-go: GTAAATATAG TTTAACCAAA ACATCAGATT GTGAATCTGA  
TAACAGAGGC TCACAACCCC TTATTTACC

4-or: GTAAATATAG TTTAACCAAA ACATTAGATT GTGAATCTAA  
TAATAGGGCC CCACAACCCC TTATTTACC

5- gi: GTAAACATAG TTTAATCAAACATAGATT GTGAATCTAA  
CAATAGAGGC TCGAAACCTC TTGCTTACC

Xác định quan hệ giữa các đại diện trong họ linh trưởng?

#### Bước 1:

So sánh các trình tự trên, tìm các vị trí nucleotid khác nhau giữa các trình tự, ta có ma trận khoảng cách di truyền thứ nhất:

	hu	ch	go	or	gi	
hu	-	0,0145	0,0434	0,1304	0,1739	Tần số
ch	1	-	0,0290	0,1159	0,1594	
go	3	2	-	0,0869	0,1594	
or	9	8	6	-	0,1594	
gi	12	11	11	11	-	

Các số liệu ở ma trận khoảng cách thứ nhất, khoảng cách di truyền giữa người và tinh tinh  $d(hu-ch) = 0,0145$  là nhỏ nhất. Do đó, cụm thứ nhất được chọn ( $hu-ch$ ).

#### Bước 2:

Tính khoảng cách di truyền giữa các đối tượng còn lại với cụm thứ nhất:

$$d(hu - ch)_{go} = \frac{1}{2} d(hu - go) + d(ch - go) = 0,0362$$

$$d(hu - ch)or = \frac{1}{2} d(hu - or) + d(ch - or) = 0,1231$$

$$d(hu - ch)gi = \frac{1}{2} d(hu - gi) + d(ch - gi) = 0,1667$$

Từ kết quả tính toán, thu được ma trận khoảng cách di truyền thứ hai:

	<i>hu-ch</i>	<i>go</i>	<i>or</i>	<i>gi</i>
<i>hu-ch</i>	-	0,0362	0,1231	0,1667
<i>go</i>		-	0,0869	0,1594
<i>or</i>			-	0,1594
<i>gi</i>				-

### Bước 3:

Số liệu ở ma trận khoảng cách thứ ba có  $d(hu-ch)go = 0,0362$  nhỏ nhất tạo nên cụm thứ hai là  $(hu-ch)go$ .

Tính khoảng cách di truyền các đối tượng còn lại với cụm thứ hai, ta có ma trận khoảng cách thứ ba:

$$d(hu - ch, go)or = \frac{1}{2} [ d(hu - ch)or + d(go - or) ] = 0,1050$$

$$d(hu - ch, go)gi = \frac{1}{2} d[ (hu - ch)gi + d(go - gi) ] = 0,1632$$

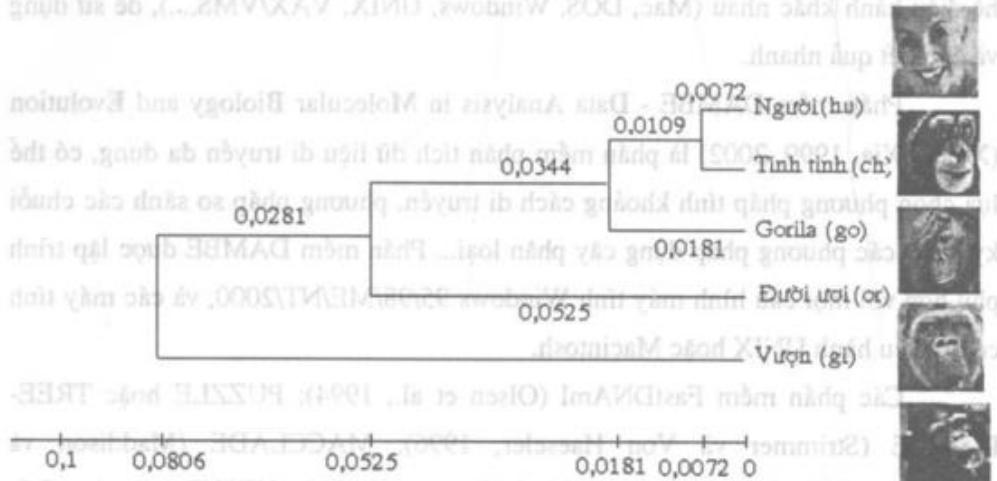
	<i>(hu-ch-go)</i>	<i>or</i>	<i>gi</i>
<i>(hu-ch-go)</i>	-	0,1050	0,1632
<i>or</i>		-	0,1594
<i>gi</i>			-

Chỉ số khoảng cách di truyền  $d(hu-ch-go)or = 0,1050$  thấp nhất tạo nên cụm thứ ba  $(hu-ch-go-or)$ , ta có:

$$d(hu - ch - go, or)gi = \frac{1}{2} [ d(hu - ch - go)gi + d(go - gi) ] = 0,1613$$

#### Bước 4:

Từ các số liệu trên dùng được cây phân loại quan hệ họ hàng giữa người, tinh tinh, gorila, dười ươi và vượn.



Hình 3.8. Cây phân loại giữa các đại diện của họ linh trưởng

Số liệu khoảng cách di truyền biểu hiện ở cây phân loại cho thấy, trong các đại diện của họ linh trưởng được nghiên cứu, người có quan hệ gần gũi nhất với tinh tinh và khác xa nhất với vượn.

#### 4. Sử dụng các phần mềm trong phân tích số liệu và dựng cây phân loại

Các phần mềm chuyên dụng phân tích dữ liệu di truyền và dựng cây phân loại trên máy tính, ngày càng được sử dụng rộng rãi trong các nghiên cứu sinh học, công nghệ sinh học, nông nghiệp, y học... Phần mềm xử lý dữ liệu di truyền ngày càng nhiều, đã có hơn 300 phần mềm khác nhau được thương mại hóa. Lựa chọn phần mềm thích hợp với cấu trúc dữ liệu, cho kết quả phân tích chính xác hơn. Một

số phần mềm xử lí dữ liệu di truyền được ưa chuộng và sử dụng rộng rãi trên toàn cầu như PHYLIP, FastDNAml, DAMBE, PUZZLE, MACCLADE, NTSYSpc...

**PHYLIP - phylogeny Inference package** (Felsenstein, 1993, 1996) là phần mềm gồm khoảng 30 chương trình phân tích tiến hoá phân tử, từ tìm và so sánh các trình tự đến dựng cây phân loại. PHYLIP phù hợp với mọi cấu hình máy tính và các hệ điều hành khác nhau (Mac, DOS, Windows, UNIX, VAX/VMS,...), dễ sử dụng và cho kết quả nhanh.

**Phần mềm DAMBE - Data Analysis in Molecular Biology and Evolution** (Xuhua Xia, 1999, 2002) là phần mềm phân tích dữ liệu di truyền đa dạng, có thể lựa chọn phương pháp tính khoảng cách di truyền, phương pháp so sánh các chuỗi ký tự và các phương pháp dựng cây phân loại... Phần mềm DAMBE được lập trình phù hợp với mọi cấu hình máy tính Windows 95/98/ME/NT/2000, và các máy tính có hệ điều hành UNIX hoặc Macintosh.

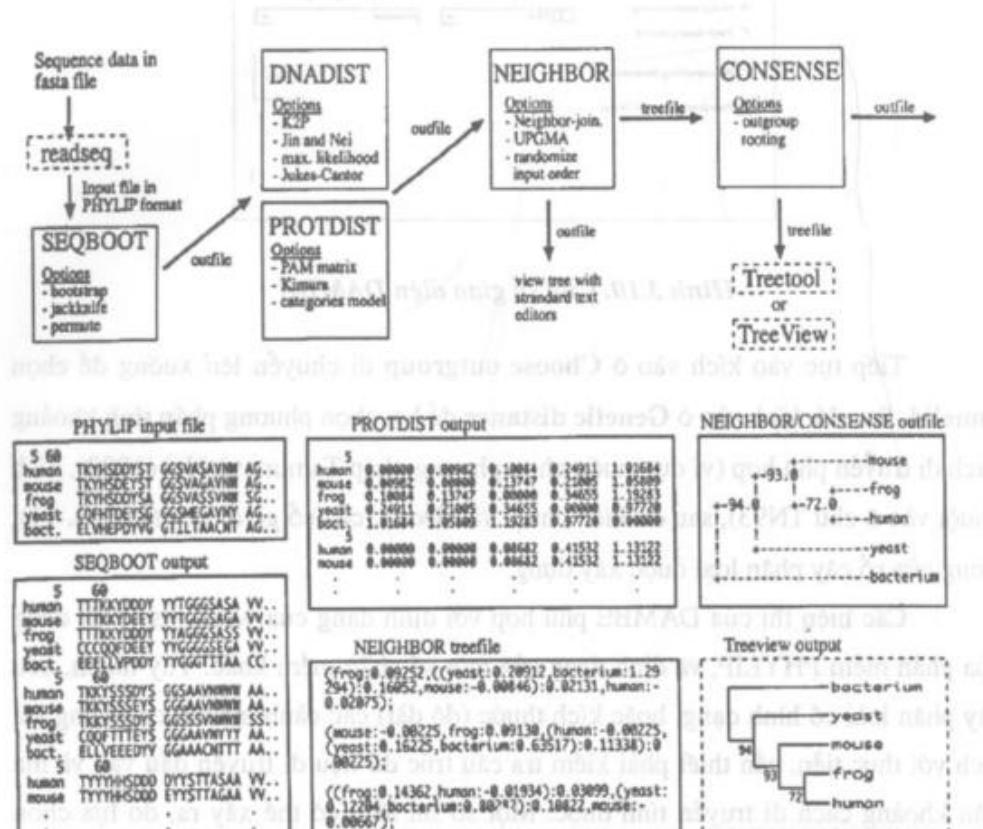
Các phần mềm FastDNAml (Olsen et al., 1994); PUZZLE hoặc TREE-PUZZLE (Strimmer và Von Haeseler, 1996); MACCLADE (Maddison và Maddison, 1992); MOLPHY (Adachi và Hasegawa, 1996); NTSYSpc version 2.0; TREE DRAWING gồm TREETOOL (X-windows), TREEDRAW (Macintosh), PHYLOGENDRON (Macintosh); TREEVIEW (Macintosh, Microsoft Windows)... là những phần mềm phân tích dữ liệu di truyền có nhiều điểm đặc trưng, phù hợp với các cấu trúc dữ liệu khác nhau.

#### Ví dụ 1: Sử dụng phần mềm PHYLIP.

Đầu tiên cần khởi động phần mềm và cẩn cứ vào cấu trúc dữ liệu là trình tự nucleotid hoặc trình tự acid amin để chọn DNADIST hay PROTDIST.

Tiếp theo, cần lựa chọn phương pháp tính khoảng cách di truyền và xử lí dữ liệu (ví dụ, chọn Nei hoặc Jukes- Cantor...). Khi đã có các số liệu về khoảng cách di truyền, tiếp tục lựa chọn phương pháp dựng cây phân loại (ví dụ, chọn UPGMA hoặc NJ), và lựa chọn dạng cây có gốc hay không gốc...

Trong mỗi bước có thể in dữ liệu đã xử lí, để so sánh làm cơ sở để lựa chọn bước tiếp theo thích hợp (hình 3.9)



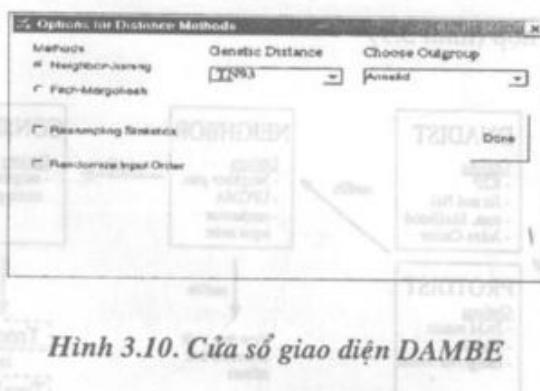
Hình 3.9. Sơ đồ phân tích dữ liệu di truyền bằng phần mềm PHYLIP

#### Ví dụ 2: Sử dụng phần mềm DAMBE

Khởi động phần mềm và kích chuột vào ô chữ cần thiết: **Phylogenetics/Distance methods/Nucleotide sequences.**

Ví dụ, muốn chọn phương pháp khoảng cách di truyền ta kích chuột vào dòng chữ **Distance methods**, cửa sổ giao diện xuất hiện (hình 3.10).

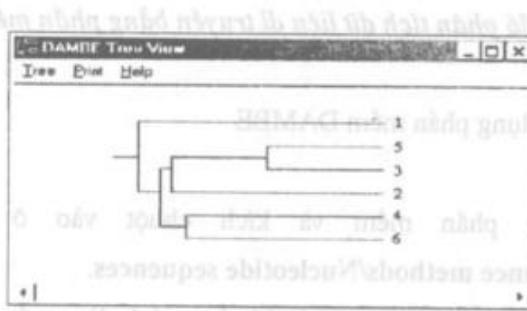
như sau: mở file tên là file định dạng là .dnd và click chuột phải vào nó và chọn Import.



Hình 3.10. Cửa sổ giao diện DAMBE

Tiếp tục vào kích vào ô Choose outgroup di chuyển lên xuống để chọn annelid. Sau đó, kích vào ô Genetic distance để lựa chọn phương pháp tính khoảng cách di truyền phù hợp (ví dụ, muốn chọn phương pháp Tamura và Nei (1993), kích chuột vào ô chữ TN93), sau đó kích chuột vào Done, cửa sổ giao diện (hình 3. 11), trong cửa sổ cây phân loại được xây dựng.

Các hiện thị của DAMBE phù hợp với định dạng của windows, định dạng của phần mềm PHYLIP, và định dạng của một số phần mềm khác. Tuy nhiên, nếu cây phân loài có hình dạng, hoặc kích thước (độ dài) các cành, nhánh có những sai lệch với thực tiễn, cần thiết phải kiểm tra cấu trúc dữ liệu di truyền đầu vào và ma trận khoảng cách di truyền tính được. Một số sai lệch có thể xảy ra, do lựa chọn phương pháp phân tích không phù hợp với cấu trúc của dữ liệu di truyền.



Hình 3.11. Cửa sổ giao diện hiện thị cây phân loại phần mềm DAMBE

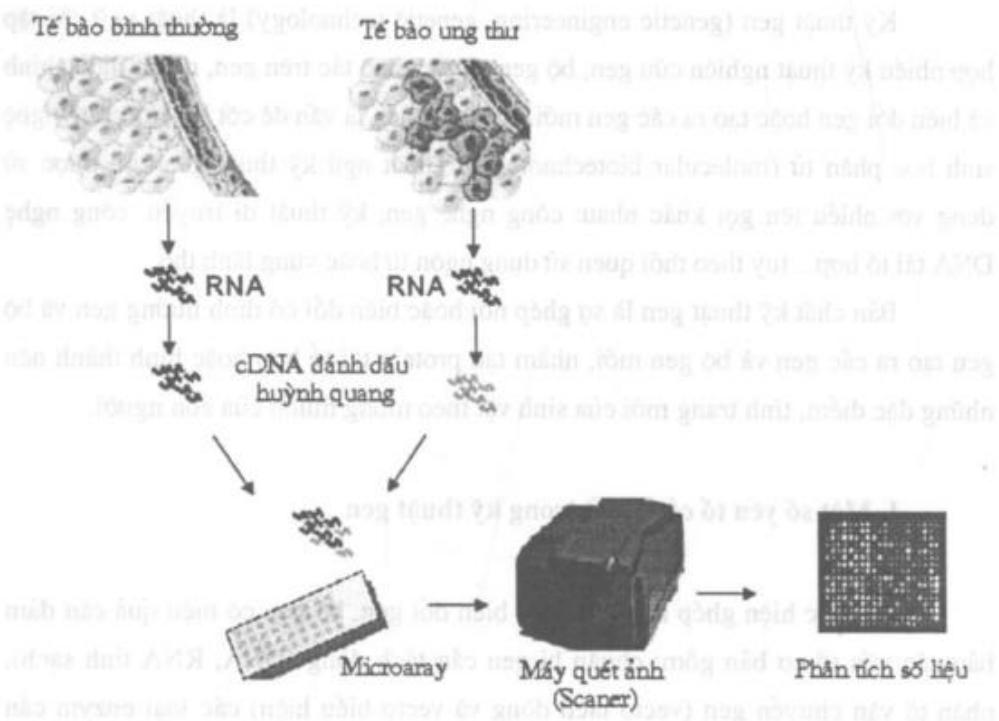
**Tài liệu tham khảo chính**

1. Cavalli-Sforza L. L., Edwards A. W. F., (1967). Phylogenetic analysis: models and estimation procedures. Am. J. Hum. Gen. 19: 233-257.
2. Edwards A. W. F., Hammond H. A., Jin L., Caskey C. T., and Chakraborty, R., (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241-253.
3. Goldman N., (2001). Molecular phylogenetics: State-of-the-art methods for looking into the past, *Trends in Genetics* 17, 262–272.6.
4. Goldstein D. B., Linaresn A. R., Cavalli-Sforza L. L and M. W. Feldman M. W., (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. PNAS USA 92: 6723-6727.
5. Graur and Li, (2000). Fundamentals of molecular evolution (2nd edition). Sinauer Assoc.
6. Lerat E., Daubin V., Moran N. A., (2003). From Gene Trees to Organismal Phylogeny in Prokaryotes: The Case of the c-Proteobacteria. PLoS Biology 1: 101-109.
7. Машуров А. М., Тхань Х. Х., Зуй Н. С, Анализ происхождения вьетнамской молочной породы скота по результатам изучения маркерных признаков. Генетика, 1992, т. 28, № 6, с. 109-122.
8. А. М. Машуров, Н. О. Сухова, Р. О. Царев, Х.Х. Тхань. Алгоритмы иммунобиохимической генетики. Новосибирск, 1998, 112 с.
9. Nei M., (1972). Genetic distance between populations. Am. Nat. 106: 283-292.
10. Nei M., (1984). Molecular Evolutionary Genetics. Columbia University Press/ New York.
11. Reynolds J., Weir B. S., Cockerham C. C., (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. Genetics 105: 767-779.
12. Rubinstein D. C., Amos W., Leggo J., Goodburn S., Jain S., Li S.H., Margolis R.L., Ross C. A., and Ferguson-Smith M. A., (1995). Microsatellite evolution - 113.190.1.105 downloaded 60930.pdf at Sun Jul 01 23:28:18 ICT 2012

- evidence for directionality and variation in rate between species. *Nature Genet.* 10: 337-343.
13. Sokal R. R. and Rohlf F. J., (1981). *Biometry*. W. H. Freeman & Ltd/ New York.
14. Weir Bruce S., (1996). *Genetic Data Analysis II: Methods for discrete population genetic data* (2nd. ed.). Sinauer Assoc., Sunderland, MA.
15. Xia X. and Xie Z., (2001). DAMBE: Data analysis in molecular biology and evolution. *Journal of Heredity* 92: 371-373
16. <http://evolution.genetics.washington.edu/phylip.html>
17. <http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/>

## PHẦN II

### NGUYÊN LÝ VÀ PHƯƠNG PHÁP NGHIÊN CỨU CƠ BẢN CỦA KỸ THUẬT GEN



#### Chương 4

##### *Nguyên lý kỹ thuật gen*

#### Chương 5

##### *Phương pháp PCR và một số kỹ thuật cơ bản*

*trong công nghệ sinh học phân tử*

## Chương 4

### NGUYÊN LÝ KỸ THUẬT GEN

Kỹ thuật gen (genetic engineering, genetic technology) là thuật ngữ chỉ tập hợp nhiều kỹ thuật nghiên cứu gen, bộ gen và các thao tác trên gen, nhằm điều chỉnh và biến đổi gen hoặc tạo ra các gen mới. Kỹ thuật gen là vấn đề cốt lõi của công nghệ sinh học phân tử (molecular biotechnology). Thuật ngữ kỹ thuật gen còn được sử dụng với nhiều tên gọi khác nhau: công nghệ gen, kỹ thuật di truyền, công nghệ DNA tái tổ hợp... tùy theo thói quen sử dụng ngôn từ hoặc vùng lãnh thổ.

Bản chất kỹ thuật gen là sự ghép nối hoặc biến đổi có định hướng gen và bộ gen tạo ra các gen và bộ gen mới, nhằm tạo protein tái tổ hợp hoặc hình thành nên những đặc điểm, tính trạng mới của sinh vật theo mong muốn của con người.

#### I. Một số yếu tố cần thiết trong kỹ thuật gen

Để thực hiện ghép nối gen hoặc biến đổi gen, bộ gen có hiệu quả cần đảm bảo các yếu tố cơ bản gồm: chuẩn bị gen cần tách dòng (DNA, RNA tinh sạch), nhân tố vận chuyển gen (vectơ tách dòng và vectơ biểu hiện) các loại enzym cần thiết, hệ thống tế bào chủ... ngoài ra, cần đòi hỏi một số thiết bị, hóa chất có chất lượng cao.

##### *I. Chuẩn bị gen cần tách dòng*

Gen cần tách dòng còn được gọi với nhiều thuật ngữ khác nhau: gen cần

thiết, đoạn xen DNA, DNA insert... Gen cần tách dòng thường là những gen mã hoá các protein, enzym hoặc gen mã hoá các đặc điểm quý (năng suất cao, khả năng chống bệnh...) cần thiết cho con người trong thực tiễn sản xuất hoặc trong nghiên cứu.

Gen cần tách dòng có thể được chuẩn bị theo hai con đường: tách chiết từ bộ gen vi sinh vật, động thực vật và con người hoặc tổng hợp nhân tạo các oligonucleotid. Trong thực tế, các gen mang các đặc điểm quý hiếm, cần thiết cho con người chủ yếu được tách chiết từ bộ gen vi sinh vật hoặc từ bộ gen sinh vật bậc cao.

### 1. 1. Tách chiết DNA tổng số

Khi xác định một loài vi sinh vật hoặc loài sinh vật bậc cao có các gen mang các đặc điểm cần thiết, người ta có thể sử dụng các kỹ thuật khác nhau để tách chiết DNA bộ gen, từ đó phân lập các gen cần thiết.

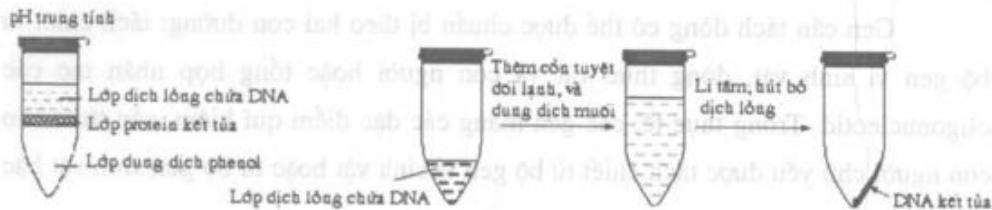
Tách chiết DNA bộ gen (DNA genome) có thể thực hiện theo các bước chủ yếu sau:

Lấy mẫu sinh học (mô tế bào, lông tóc, dịch sinh học...) theo các phương pháp thông thường. Phá vỡ cấu trúc tế bào bằng enzym lysozym hoặc xử lí mẫu với các chất tẩy mạnh như SDS (sodium dodecyl sulfat), EDTA hoặc bằng phương pháp siêu âm, đun sôi...

Gây tủa protein bằng dung dịch phenol - chlorophor và enzym proteinase K, sau khi li tâm trong ống nghiệm hình thành 3 lớp dung dịch: lớp dung dịch phenol nằm dưới đáy ống nghiệm, lớp tủa protein nằm ở giữa, trên cùng là dung dịch lỏng chứa DNA.

Hút lớp dung dịch lỏng chứa DNA ở phía trên sang ống eppendorf sạch, bổ sung sodium acetate 0,3 M. Thêm 2,5 thể tích cồn tuyệt đối (100%) ở - 20°C để gây tủa DNA, li tâm thu tủa DNA.

Để tinh sạch DNA cần rửa tủa DNA một vài lần bằng cồn 70%, làm khô và hòa tan trong dung dịch TE (10 mM TRIS pH 7,5 + 1 mM EDTA) hoặc nước khử ion để tiếp tục nghiên cứu.



Hình 4.1. Sơ đồ tách chiết DNA tổng số

Kiểm tra độ sạch DNA bằng do quang phổ kế, xác định tỉ lệ  $OD_{260}/OD_{280}$  hoặc bằng phương pháp điện di. DNA tinh sạch thu được bảo quản ở  $-20^{\circ}\text{C}$ , có thể giữ được trong vòng một vài tháng.

### 1.2. Tách chiết DNA plasmid

Chuẩn bị dịch nuôi cấy vi khuẩn thích hợp, thường sử dụng môi trường LB có bổ sung chất kháng sinh để vi khuẩn có số lượng plasmid cao nhất. Nuôi vi khuẩn ở nhiệt độ  $37^{\circ}\text{C}$  trên máy lắc với tốc độ 200 vòng/ phút qua đêm, li tâm với tốc độ 5 000 vòng/ phút trong 5 - 6 phút thu sinh khối (cận tế bào).

Phá vỡ tế bào bằng hoá chất hoặc siêu âm. Sau đó xử lí natri acetat ( $\text{pH} = 4,3$ ) và li tâm 12000 vòng trong 10 phút. Sau li tâm, DNA bộ gen và protein kết tủa ở đáy ống, lớp dịch nổi phía trên chứa DNA plasmid. Thu lớp dịch nổi chứa plasmid chuyển sang ống khác, thêm ethanol 100% (hoặc isopropanol) lạnh để kết tủa DNA plasmid.

Li tâm 14000 vòng trong 10 - 15 phút, thu tủa DNA plasmid hoà trong nước khử ion hoặc dung dịch TE, sau đó kiểm tra độ tinh sạch bằng điện di hoặc quang phổ kế.

### 1. 3. Tách chiết RNA

#### - *Tách chiết RNA thông tin*

Tách chiết RNA thông tin (mRNA) thường sử dụng các mồi oligo - dT. Gồm các bước cơ bản sau: Mẫu (1g lá non, 0,5 - 1g thịt, gan) để trong tủ - 85°C trong 2 - 3 giờ, sau đó nghiền trong nitơ lỏng thành bột mịn. Thêm 15 ml dung dịch đệm có thành phần 200 mM NaCl; 200 mM Tris pH = 7,5; 1,5 mM MgCl<sub>2</sub> và 2% SDS, ủ 60 phút ở 45°C. Li tâm 6 000 vòng trong 5 - 10 phút. Thêm NaCl 0,5 M, trộn kỹ. Sau đó thêm mồi oligo - dT (các đoạn oligonucleotid dTTP ngắn), lắc nhẹ 1 giờ và li tâm ở tốc độ 3 000 vòng trong 5 phút. Rửa oligo - dT với đệm thích hợp, li tâm 3 000 vòng trong 5 phút sau đó rửa trong DEPC (Diethyl pyrocarbonate), li tâm để thu mRNA.

#### - *Tách chiết và tinh sạch RNA tổng số bằng mini kit QIA gen*

Để tách chiết các mRNA, rRNA 5 S và rRNA 5,8 S có kích thước nhỏ hơn 200 nucleotid, người ta thường sử dụng mini kit của hãng QIA gen. Các bước chủ yếu gồm: Phá tế bào bằng 450 µl dung dịch đệm, li tâm 15 000 vòng / 2 phút, thu dịch nổi phía trên. Thêm 225 µl ethanol (96 - 100%) trộn đều bằng pipet. Cho tất cả mẫu qua cột silicagel, li tâm 10 000 vòng trong 15 giây, loại bỏ phần nước. Rửa cột để loại bỏ tạp chất (rửa 3 lần với đệm RPE), sau đó đưa các cột vào li tâm với tốc độ 14 000 vòng trong 1 phút và chuyển cột sang ống eppendorf 1,5 ml. Hoà tan RNA trong cột bằng 50 µl nước tinh sạch (không chứa RNase). Li tâm 10 000 đến 13 000 vòng trong 1 - 2 phút thu được RNA tinh sạch.

#### - *Tách chiết RNA tổng số từ máu bằng kit của hãng Invitrogen*

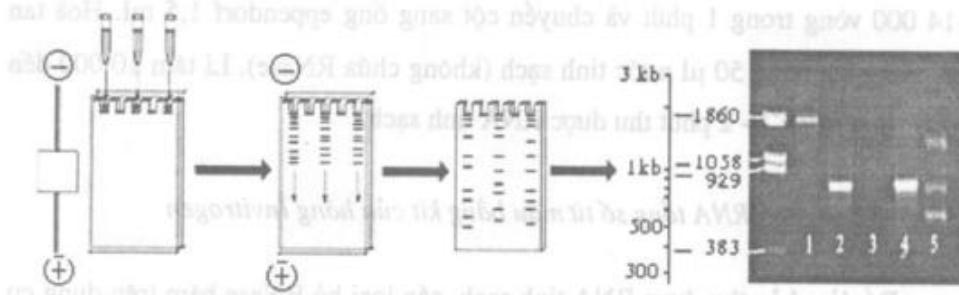
Để đảm bảo thu được RNA tinh sạch, cần loại bỏ RNase bám trên dụng cụ thiết bị bằng cách ngâm dụng cụ trong dung dịch DEPC 1% trong 8 - 12 giờ, hấp

khử trùng với nhiệt độ và áp suất cao. Lấy 200 µl máu cho vào ống eppendorf, bổ sung proteinase K, lắc nhẹ 30 giây, ủ 20 phút ở 37°C. Li tâm 6 phút tốc độ 12 000 vòng / phút thu dịch, thêm 400 µl isopropanol và trộn đều. Chuyển dịch sang cột S.N.A.P. Li tâm 1 phút tốc độ 12000 vòng / phút, bỏ dịch li tâm và rửa cột 2 lần bằng dung dịch rửa 1X, trong 2 phút.

Sử dụng 120 µl nước khử ion để thu axit nucleic tổng số. Loại DNA bằng cách cho thêm 15 µl đậm 10X cho DNase, ủ 37°C trong 8 - 10 phút, thêm 300 µl Isopropanol và đảo ống 4 - 5 lần. Chuyển dịch vào một cột S.N.A.P mới. Li tâm trong 1 - 2 phút ở tốc độ 12 000 vòng / phút, bỏ dịch li tâm và rửa cột 2 lần bằng dung dịch rửa 1X. Sử dụng 120 µl nước khử ion để thu RNA tổng số tinh sạch.

#### 1.4. Kiểm tra độ tinh sạch DNA, RNA bằng điện di trên gel agarose

Chuẩn bị đậm điện di TAE × 1 hoặc TBE × 1 và đổ bản gel với thành phần (0,4 - 0,8 g agarose; 2ml TAE × 50; 98 ml H<sub>2</sub>O), đun sôi cách thuỷ hoặc đặt vào lò vi sóng cho tan hoàn toàn, đổ vào khuôn. Đặt bản gel vào hộp điện di đổ dung dịch đậm ngập bản gel. Chấm mẫu vào các giếng trên bản gel. Chạy điện di 90 - 100V trong 30 - 40 phút, lấy bản gel nhuộm EtBr. Rửa bản gel và hiện hình DNA để kiểm tra kết quả điện di. Kết quả điện di cho các vạch gọn, phân tách rõ ràng nhã là DNA trong mẫu đảm bảo độ tinh sạch cho nghiên cứu. Nếu các vạch kéo dài hoặc phân tách không rõ cần xử lí lại từ khâu xử lí proteinase K, để đảm bảo độ tinh sạch của mẫu nghiên cứu.



Hình 4.2. Sơ đồ kỹ thuật điện di tinh sạch DNA

### 1. 5. Kiểm tra độ tinh sạch DNA, RNA bằng quang phổ kế

Trên cơ sở mức độ hấp thụ ánh sáng khác nhau của các bazơ nitơ trong các mạch đơn DNA, DNA mạch kép và RNA có thể xác định hàm lượng của DNA, RNA trong dịch chiết bằng các chỉ số hấp thụ quang phổ.

Kiểm tra độ tinh sạch của DNA trong dịch chiết dựa vào tỉ lệ  $A_{260} / A_{280}$ . Giá trị mật độ quang ở bước sóng 260 nm ( $OD_{260\text{ nm}}$  - Optical Density) của các mẫu DNA cho phép xác định nồng độ DNA trong dung dịch. Protein có phổ hấp thụ cao nhất ở bước sóng 280 nm, đồng thời hấp thụ ở bước sóng 260 nm, do vậy tỉ lệ  $A_{260} / A_{280}$  biểu thị mức độ protein còn sót trong dịch chiết. DNA (RNA) được coi là tinh sạch khi giá trị  $A_{260\text{ nm}} / A_{280\text{ nm}} = 1,8 - 2,0$ .

Chỉ số đo quang phổ của dịch chiết DNA (RNA) ngoài kiểm tra độ tinh sạch, còn dùng để xác định nồng độ DNA (RNA). Hàm lượng DNA (RNA) trong dịch chiết được tính theo các biểu thức sau:

$$A_{260\text{ nm}} = 1,0 = 50 \mu\text{g} / \text{ml DNA} \text{ sợi đôi}$$

$$A_{260\text{ nm}} = 1,0 = 37 \mu\text{g} / \text{ml DNA} \text{ sợi đơn}$$

$$A_{260\text{ nm}} = 1,0 = 40 \mu\text{g} / \text{ml RNA}$$

Tính nồng độ DNA (RNA) ngoài chỉ số OD, cần lưu ý đến hệ số pha loãng của dịch chiết. Hàm lượng DNA trong dịch chiết được tính bằng biểu thức:

$$C_{DNA} (\mu\text{g} / \text{ml}) = A_{260\text{ nm}} \times 50 \times (\text{độ pha loãng})$$

Ví dụ, dịch chiết DNA bộ gen tế bào vi khuẩn, sau khi pha loãng 20 lần có giá trị mật độ quang đo được  $A_{260\text{ nm}} = 0,6$ . Có thể xác định được hàm lượng DNA trong dịch chiết:

$$C_{DNA} = 0,6 \times 50 \times 20 = 600 \mu\text{g} / \text{ml.}$$

Hàm lượng DNA (giá trị  $C_{DNA}$ ) được xác định chính xác khi dịch chiết DNA tinh sạch hoàn toàn, do đó cần kiểm tra độ tinh sạch của dịch chiết trước khi xác định nồng độ DNA.

Sau khi tách chiết và tinh sạch DNA (hoặc RNA), có thể chuẩn bị gen cần tách dòng bằng nhiều cách: sử dụng các loại enzym giới hạn kết hợp với phương pháp điện di với thang DNA chuẩn để lựa chọn các đoạn DNA theo kích thước đã

biết hoặc sử dụng kỹ thuật PCR với các mồi đặc hiệu để nhân (khuếch đại) chọn lọc các gen hoặc các đoạn DNA cần thiết... (chương 5)

## **2. Enzym giới hạn**

### **2.1. Khái niệm enzym giới hạn**

Hiện tượng giới hạn và enzym giới hạn do Hamilton Smith và cộng sự phát hiện đầu tiên (1970) ở vi khuẩn *Hemophilus influenzae*. Bình thường tế bào vi khuẩn bị nhiễm phage thường bị phage phá huỷ. Một số chủng vi khuẩn khi bị nhiễm phage, không bị phage phá huỷ mà có khả năng phân huỷ DNA phage, nhờ một số loại enzym đặc hiệu trong tế bào vi khuẩn. Các loại enzym trong tế bào vi khuẩn có thể cắt DNA phage ở những vị trí nhất định thành những đoạn ngắn, những enzym này được gọi là enzym giới hạn.

Enzym giới hạn (Restriction Enzyme - RE) là các enzym thuộc nhóm enzym endonuclease, có khả năng nhận biết các trình tự nucleotid đặc hiệu trên phân tử DNA (khoảng 4 - 8 nucleotid), cắt đặc hiệu phân tử DNA ở những vị trí nhất định thành những đoạn ngắn.

Căn cứ vào khả năng nhận biết và vị trí cắt đặc hiệu DNA, người ta chia enzym giới hạn làm ba nhóm:

Nhóm thứ nhất gồm các enzym giới hạn nhận biết trình tự nucleotid đặc hiệu, sau đó trượt lên phía trước cắt phân tử DNA ở cách vị trí nhận biết khoảng 1 000 - 5 000 nucleotid.

Nhóm thứ hai gồm các enzym giới hạn nhận biết các trình tự nucleotid đặc hiệu và cắt phân tử DNA ngay tại vị trí nhận biết. Các enzym nhóm này có vai trò quan trọng trong kỹ thuật gen.

Nhóm thứ ba gồm các enzym giới hạn nhận các trình tự nucleotid đặc hiệu và cắt phân tử DNA ở cách vị trí nhận biết khoảng 20 nucleotid về phía trước.

Do vị trí cắt của các enzym giới hạn nhóm 1 và nhóm 3 không được biết chính xác nên ít được sử dụng trong kỹ thuật gen. Các enzym giới hạn sử dụng trong

kỹ thuật gen và được thương mại hóa hiện nay chủ yếu là các enzym giới hạn thuộc nhóm thứ hai.

Mỗi loại enzym giới hạn có hoạt tính cao trong các dung dịch đậm phù hợp và ở những điều kiện nhất định về nhiệt độ, độ pH... (Ví dụ, sử dụng enzym EcoR I để cắt DNA của tế bào động vật có thể sử dụng dung dịch đậm gồm 100 mM Tris - HCl pH = 7,5; 5 mM MgCl<sub>2</sub>; 100 mg BSA / ml; 0,15% TritonX-100). Phản ứng của enzym giới hạn thực hiện trong một thể tích càng nhỏ càng tốt, đảm bảo cho enzym giới hạn tiếp xúc tốt nhất với cơ chất, có hiệu quả cao nhất. Thể tích enzym giới hạn bằng 1/10 thể tích phản ứng là thích hợp nhất.

Enzym giới hạn không mang tính đặc hiệu loài, nghĩa là một loại enzym giới hạn được tách chiết từ vi khuẩn nhận biết trình tự đặc hiệu gồm 6 cặp nucleotid, có thể nhận biết và cắt các phân tử DNA của tế bào động vật, thực vật và vi sinh vật ở trình tự 6 cặp nucleotid giống như ở vi khuẩn.

Vị trí nhận biết và cắt đặc hiệu các nucleotid trên phân tử DNA ở mỗi loại enzym giới hạn khác nhau, gọi là vị trí giới hạn. Các đoạn DNA có kích thước nhỏ tạo nên do enzym giới hạn cắt phân tử DNA gọi là đoạn giới hạn. Số lượng, kích thước đoạn giới hạn phụ thuộc vào số vị trí giới hạn trên phân tử DNA và đặc điểm nhận biết đặc trưng của mỗi loại enzym giới hạn.

## 2. 2. Tên gọi của enzym giới hạn

Tên gọi các enzym giới hạn được thống nhất kí hiệu bằng 3 - 5 chữ cái, kèm theo các số La Mã. Ví dụ:

Chữ cái đầu tiên viết chữ in hoa, chỉ tên chi của vi khuẩn (hoặc sinh vật) đã tách chiết được enzym giới hạn.

Hai chữ cái tiếp theo viết chữ in thường, chỉ loài hoặc giống của sinh vật đã tách chiết được enzym giới hạn.

Chữ cái tiếp theo viết hoa chỉ tên chủng vi khuẩn đã tách chiết được enzym giới hạn.

Chữ số La Mã chỉ số thứ tự dòng vi khuẩn hoặc sinh vật, mà enzym giới hạn được phát hiện.

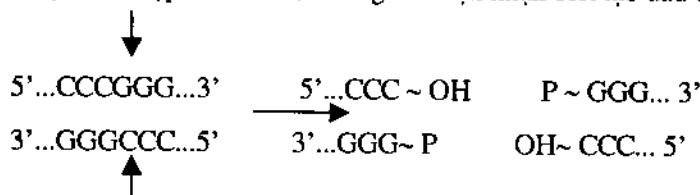
Ví dụ:

Enzym	Chi,	Loài (Giống)	Chủng	Thứ tự
<b>Eco RI</b>	E	co	R	I
	<i>Escherichia</i>	coli	Ry13	
<b>Hae III</b>	H	ae		III
	<i>Hemophilus</i>	<i>aegypticus</i>		
<b>Hind III</b>	H	in	d	III
	<i>Hemophilus</i>	<i>influenzae</i>		

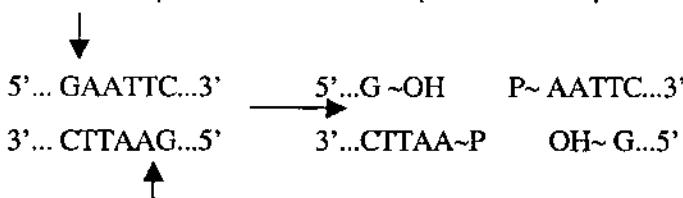
### 2. 3. Các kiểu cắt của enzym giới hạn

Mỗi enzym giới hạn nhận biết và cắt đặc hiệu ở một trình tự nucleotid trên phân tử DNA. Nếu enzym giới hạn nhận biết trình tự DNA đặc hiệu gồm bốn cặp nucleotid (4 bp), nghĩa là trên phân tử DNA cứ  $4^4$  (256) cặp nucleotid có một vị trí cắt. Khi các enzym giới hạn nhận biết trình tự DNA đặc hiệu có 6 cặp nucleotid (6 bp), thì cứ  $4^6$  (4096) cặp nucleotid có vị trí cắt.

Tùy đặc điểm của từng loại enzym giới hạn, đoạn DNA cắt có đầu bằng hoặc đầu đầu so le. Enzym giới hạn cắt cả 2 mạch DNA cùng một điểm tạo các đoạn cắt có đầu bằng (blunt ends). Các đoạn DNA cắt đầu bằng không có khả năng tự dính lại với nhau, để nối các đoạn DNA cắt đầu bằng cần sử dụng enzym nối (DNA ligase) hoặc các adaptor chuyên dụng cho mỗi loại enzym. Ví dụ, enzym *Sma* I nhận biết đoạn đặc hiệu có 6 cặp nucleotid, cắt ở giữa đoạn nhận biết tạo đầu bằng:



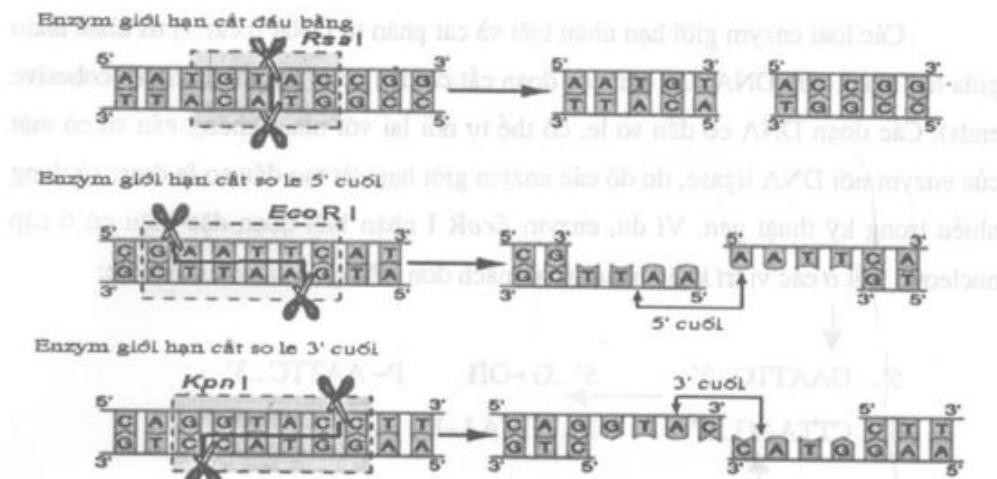
Các loại enzym giới hạn nhận biết và cắt phân tử DNA ở các vị trí khác nhau giữa hai mạch đơn DNA, tạo nên các đoạn cắt có đầu so le (hay đầu dính - cohesive ends). Các đoạn DNA có đầu so le, có thể tự nối lại với nhau không cần sự có mặt của enzym nối DNA ligase, do đó các enzym giới hạn cắt tạo đầu so le được sử dụng nhiều trong kỹ thuật gen. Ví dụ, enzym *EcoR I* nhận biết đoạn đặc hiệu có 6 cặp nucleotid, cắt ở các vị trí khác nhau ở hai mạch đơn DNA tạo các đầu cắt sole:



Bảng 4. 1. Một số loại enzym giới hạn và các vị trí cắt đặc hiệu

Enzym	Vị trí cắt	bp	Đầu cắt	Nguồn gốc
<i>EcoRI</i>	G/AATTC	6	5' đầu cuối	<i>Escherichia coli RY13</i>
<i>Bam HI</i>	G/GATCC	6	5' đầu cuối	<i>Bacillus amyloliquefaciens H</i>
<i>Bgl II</i>	A/GATCT	6	5' đầu cuối	<i>Bacillus globigii</i>
<i>PstI</i>	CTGCA/G	6	5' đầu cuối	<i>Providencia stuartii</i>
<i>XbaI</i>	C/CCGGG	6	5' đầu cuối	<i>Xanthomonas malvacearum</i>
<i>SmaI</i>	CCC/GGG	6	Đầu bằng	<i>Serratia marcescens</i>
<i>Sau3A</i>	/GATC	4	5' đầu cuối	<i>Staphylococcus aureus 3A</i>
<i>AluI</i>	AG/CT	4	Đầu bằng	<i>Arthrobacter luteus</i>
<i>NoI</i>	GC/GGCCGC	8	5' đầu cuối	<i>Norcardia olitidis-caviarum</i>
<i>PacI</i>	TTAAT/TAA	8	3' đầu cuối	<i>Pseudomonas alcaligenes</i>

Enzym giới hạn cắt đầu so le gồm 2 loại. Các enzym cắt tạo đầu so le 5' cuối là enzym giới hạn cắt phân tử DNA, tạo nên mạch đơn DNA có đầu 5' dài hơn mạch đơn đầu 3' (ví dụ, *EcoR I*, *BamH I*...). Các enzym cắt tạo đầu so le 3' cuối là enzym giới hạn cắt phân tử DNA, tạo nên mạch đơn DNA có đầu 3' dài hơn mạch đơn đầu 5' (ví dụ, *Kpn I*, *PacI*...).

**Hình 4.3. Các kiểu cắt của enzym giới hạn**

Nghiên cứu phát hiện các loại enzym giới hạn mới và sản xuất enzym giới hạn là một lĩnh vực có lợi ích kinh tế cao, không đòi hỏi diện tích mặt bằng lớn. Hiện nay, sản xuất enzym giới hạn là “ngành công nghiệp” được nhiều phòng thí nghiệm, nhiều hãng sản xuất và các quốc gia quan tâm.

### 3. Các loại enzym thường sử dụng trong kỹ thuật gen

Trong kỹ thuật gen, ngoài các enzym giới hạn còn sử dụng nhiều loại enzym khác nhau, trong đó các loại enzym xúc tác tổng hợp DNA, enzym nối DNA và các enzym thuỷ phân DNA, RNA có vai trò rất quan trọng.

Mỗi loại enzym có các đặc tính khác nhau, có hoạt tính cao trong những điều kiện thích hợp nhất định. Do vậy, tuỳ theo mục đích của nghiên cứu và điều kiện cụ thể của phòng thí nghiệm để lựa chọn các loại enzym phù hợp, có hiệu quả cao trong nghiên cứu.

### 3.1. Enzym xúc tác tổng hợp DNA, RNA

Nhóm enzym xúc tác tổng hợp DNA gọi chung là DNA polymerase, gồm rất nhiều loại khác nhau có vai trò xúc tác gắn các nucleotid mới vào chuỗi mạch đơn DNA đang tổng hợp. Một số loại enzym DNA polymerase điển hình là: Taq DNA polymerase, Stoffel DNA polymerase, Pfu DNA polymerase Vent DNA polymerase, T<sub>4</sub> DNA polymerase...

#### - Taq DNA polymerase

Taq DNA polymerase là loại enzym xúc tác tổng hợp DNA được ứng dụng rộng rãi để tổng hợp DNA nhân tạo bằng phản ứng PCR. Taq DNA polymerase gồm 832 axit amin, có trọng lượng phân tử khoảng 94 kD. Taq DNA polymerase được tách chiết từ vi khuẩn *Thermus aquaticus* sống ở suối nước nóng (từ 70°C - 85°C), chịu được nhiệt độ cao không mất các hoạt tính xúc tác. Taq DNA polymerase có hoạt tính exonuclease 5'-3', không có hoạt tính exonuclease 3'-5'. Taq DNA polymerase được sản xuất phổ biến và bán rộng rãi ở nhiều nước trên thế giới, ở Việt Nam một số phòng thí nghiệm đã sản xuất được Taq DNA polymerase có hoạt tính cao, ở qui mô phòng thí nghiệm (Đại học Quốc gia Hà Nội, Viện Sốt rét và Kí sinh trùng - Côn trùng Trung ương....)

Taq DNA polymerase có hoạt tính cao nhất ở nhiệt độ 75°C - 80°C, xúc tác tổng hợp DNA với tốc độ 130 - 300 nucleotid / giây. Ở các nhiệt độ thấp hoạt tính của Taq DNA polymerase giảm dần: ở nhiệt độ 70°C, Taq DNA polymerase có tốc độ tổng hợp DNA khoảng 60 nucleotid / giây; ở 55°C - 24 nucleotid / giây; ở 37°C - khoảng 1 - 2 nucleotid / giây. Hoạt tính của Taq DNA polymerase chịu ảnh hưởng rất mạnh bởi ion Mg<sup>++</sup>. Ở điều kiện thích hợp với nồng độ MgCl<sub>2</sub> bằng 2 mM, hoạt tính của Taq DNA polymerase đạt tới mức tối đa. Khi nồng độ MgCl<sub>2</sub> tăng quá cao, làm giảm hoạt tính của enzym. Nồng độ MgCl<sub>2</sub> từ 4 - 6 mM hoạt tính của Taq DNA polymerase giảm 30%, ở nồng độ 8 mM MgCl<sub>2</sub> hoạt tính giảm đi 50%. Dung dịch đậm thích hợp cho enzym Taq DNA polymerase: 50 mM Tris - HCl (pH = 9,0), 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 200 mM dNTP mỗi loại (ATP, dGTP, dCTP, dTTP).

- *Stoffel DNA polymerase*

Stoffel DNA polymerase là enzym xúc tác tổng hợp DNA, tạo nên từ sự biến đổi cấu trúc enzym Taq DNA polymerase (cắt bớt 289 aa so với enzym Taq DNA polymerase). Stoffel DNA polymerase có trọng lượng phân tử 61 kD do hãng Perkin Elmer sản xuất, không bị giảm hoạt tính exonuclease 5' - 3'. Stoffel DNA polymerase khắc phục được một số điểm hạn chế của enzym Taq DNA polymerase như bền hơn trong điều kiện nhiệt độ cao. Enzym Stoffel DNA polymerase không bị thay đổi hoạt tính khi nồng độ của Mg<sup>++</sup> biến đổi từ 2 mM đến 10 mM.

Ở nhiệt độ 97,5°C sau 20 phút hoạt tính enzym Stoffel DNA polymerase giảm 50% (enzym Taq DNA polymerase hoạt tính giảm 50% sau 5 - 6 phút). Tính chất chịu được nhiệt cao của enzym Stoffel DNA polymerase rất quan trọng trong phản ứng PCR, khi biến tính DNA ở nhiệt độ cao với thời gian lâu hơn sẽ giúp tách tốt hơn với những đoạn DNA khuôn có tỉ lệ G - C cao.

- *Pfu DNA polymerase*

Pfu DNA polymerase là enzym xúc tác tổng hợp DNA, được tách chiết từ vi khuẩn *Pyrococcus furiosus*, trọng lượng phân tử 92 kD. Enzym Pfu DNA polymerase có cả hoạt tính exonuclease 3'-5' và exonuclease 5'-3'. Nhiệt độ thích hợp của enzym Pfu DNA polymerase từ 72°C đến 78°C. Trong điều kiện nhiệt độ cao tới 95°C sau 1 giờ enzym còn khoảng 5% hoạt tính. Enzym Pfu DNA polymerase xúc tác tổng hợp DNA với tốc độ chậm khoảng 12 nucleotid / 1 giây. Tuy tốc độ tổng hợp chậm, nhưng tỉ lệ sai sót (bắt cặp sai các nucleotid) trong quá trình tổng hợp DNA nhỏ hơn Taq DNA polymerase nhiều lần, nên Pfu DNA polymerase là loại enzym xúc tác tổng hợp DNA được ưa chuộng trong kỹ thuật gen.

- *T<sub>4</sub> DNA polymerase*

Enzym T<sub>4</sub> DNA polymerase là enzym được tách chiết đầu tiên ở thực khuẩn thể T<sub>4</sub> kí sinh vi khuẩn *E. coli*. Enzym T<sub>4</sub> DNA polymerase có hoạt tính 3' - 5' mạnh do đó được sử dụng nhiều trong kỹ thuật gen.

- *Enzym phiên mã ngược*

Enzym phiên mã ngược (reverse transcriptase) là những enzym có khả năng xúc tác tổng hợp cDNA từ mạch khuôn là RNA ở các retrovirus. Enzym phiên mã ngược được tách chiết đầu tiên từ retrovirus. Hiện nay, các enzym phiên mã ngược được thương mại hóa chủ yếu do tổng hợp nhân tạo nhờ công nghệ DNA tái tổ hợp.

- Enzym xúc tác tổng hợp RNA gồm các loại RNA polymerase thông dụng là T<sub>3</sub> RNA polymerase và T<sub>7</sub> RNA polymerase được tách chiết từ các thực khuẩn thể kí sinh vi khuẩn *E. coli*. Các enzym RNA polymerase thường có độ bền kém, khó tách chiết và tinh sạch. Enzym RNA polymerase xúc tác quá trình tổng hợp RNA từ khuôn là DNA hoặc RNA.

### 3. 2. Enzym nối DNA

Enzym DNA ligase là loại enzym xúc tác hình thành các liên kết phosphoeste nối các đoạn DNA với nhau. Mỗi loại enzym DNA ligase có các tính chất đặc trưng riêng, có khả năng nối các đoạn DNA dâu bằng hoặc dâu so le. Ví dụ, enzym *E. coli* DNA ligase chỉ nối các đoạn DNA có dâu cắt so le; còn T<sub>4</sub> DNA ligase là enzym được tách chiết từ thực khuẩn thể T<sub>4</sub>, có khả năng nối các đoạn DNA cắt dâu bằng hoặc đoạn DNA cắt sole được sử dụng rộng rãi trong kỹ thuật gen. Trong kỹ thuật gen còn sử dụng một số loại enzym khác tham gia vào phản ứng nối DNA như T<sub>4</sub> polynucleotide kinase, Alkaline photphatase...

### 3. 3. Các loại enzym phân cắt acid nucleic

Các enzym cắt phân tử DNA (DNase) hoặc cắt phân tử RNA (RNase) có vai trò quan trọng trong các kỹ thuật tinh sạch DNA, RNA, trong phản ứng phiền mã ngược tạo cDNA. Enzym cắt phân tử DNA gồm nhiều loại khác nhau: DNase I, Nuclease S1, Exonuclease.... Các loại enzym cắt RNA chủ yếu gồm RNase A, RNase H...

Ngoài các loại enzym nối trên còn rất nhiều loại enzym khác được sử dụng

trong kỹ thuật gen, tùy theo mục đích và điều kiện nghiên cứu.

#### 4. Vectơ tách dòng

##### 4. 1. Khái niệm

Vectơ tách dòng (cloning vector) là các phân tử DNA có kích thước nhỏ cho phép cài (gắn) các đoạn DNA cần thiết, có khả năng tái bản không phụ thuộc vào sự phân chia của tế bào, tồn tại trong tế bào vật chủ qua nhiều thế hệ không gây biến đổi bộ gen của tế bào vật chủ.

Vectơ tách dòng phải có nhiều điểm cắt đặc hiệu duy nhất cho các loại enzym giới hạn khác nhau, đồng thời mang các gen “tín hiệu” (gen chỉ thị màu hoặc chỉ thị kháng sinh) để dễ dàng nhận biết các tế bào mang vectơ tái tổ hợp. Vectơ tách dòng được sử dụng như một “phương tiện vận chuyển gen” nhằm tạo số lượng lớn các bản sao của một gen hoặc một trình tự DNA trong các tế bào chủ. Vectơ tách dòng có vai trò quan trọng trong tách dòng gen, giải trình tự gen, hoặc xây dựng các bản đồ gen...

##### 4. 2. Các loại vectơ tách dòng

Vectơ tách dòng gồm nhiều loại khác nhau: plasmid, phage, cosmid, phagemid, virus, nhiễm sắc thể nhân tạo vi khuẩn, nhiễm sắc thể nhân tạo nấm men... Mỗi loại vectơ tách dòng có tính chất đặc trưng riêng, cho phép cài gắn các đoạn DNA có kích thước khác nhau, tạo số lượng bản sao khác nhau (bảng 4. 2)... Trong kỹ thuật gen, căn cứ vào đặc điểm của gen cần tách dòng và mục đích nghiên cứu để lựa chọn loại vectơ tách dòng phù hợp.

###### a. Vectơ tách dòng là plasmid

Plasmid là các phân tử DNA xoắn kép dạng vòng, có kích thước nhỏ, trong tế bào chất của tế bào vi khuẩn hoặc tế bào nấm men. Kích thước trung bình của plasmid từ 1 - 5 kb, plasmid có khả năng tự tái bản độc lập với sự phân chia tế bào.

Plasmid cho cài gắn các đoạn DNA lạ, không gây ảnh hưởng tới chức năng và hoạt động của tế bào. Plasmid ở vi khuẩn mang một số ít gen như các gen xác định giới tính F, gen kháng chất kháng sinh, gen sinh độc tố... Mỗi loài vi khuẩn có các loại plasmid đặc trưng riêng. Ví dụ, vi khuẩn *E. coli* có nhiều loại plasmid khác nhau: plasmid giới tính F, plasmid *Col E1* mang gen sinh độc tố colicin úc chế sinh trưởng và tiêu diệt các loại vi khuẩn khác...

Qui ước, viết tên một plasmid gồm chữ p viết thường (viết tắt của chữ plasmid) đầu tiên, sau đó là một hoặc hai chữ tiếp theo viết hoa chỉ chữ đầu tên tác giả tạo nên plasmid hoặc tên vi khuẩn phát hiện plasmid và chữ số chỉ thứ tự chủng vì khuẩn có plasmid đó. Ví dụ, pBR322: B - Bolivar, R - Rodriguez là tên hai tác giả tạo nên plasmid này, 322 chỉ số thứ tự.

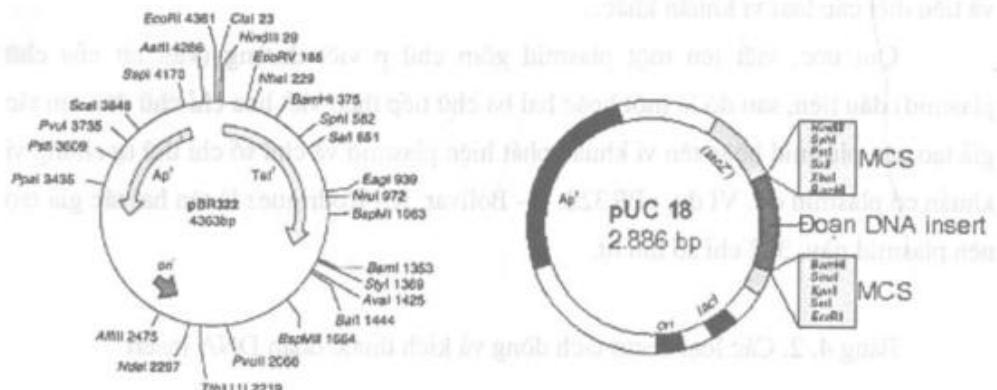
Bảng 4. 2. Các loại vectơ tách dòng và kích thước đoạn DNA insert

Loại vectơ	Kích thước DNA insert
Plasmid	1-5 kb
Phage	15-25 kb
Phagemid	20-10 kb
Cosmid	30-50 kb
BAC (thiết kế từ một phần DNA <i>E. coli</i> và các plasmid)	100-300 kb
TAC (thiết kế từ một phân đoạn vir gen của Ti-plasmid)	100-300 kb
BiBAC (thiết kế từ một phần DNA bộ gen <i>Agrobacterium</i> )	100-500 kb
PAC (bộ gen của phage P1 + plasmid )	100-300 kb
YAC (thiết kế từ một phân DNA bộ gen của nấm men)	2. 000 kb
Viral vector: Adenoviral vector, Retroviral vector, AAV...	

Plasmid được sử dụng làm vectơ tách dòng phải có các vị trí cắt đặc hiệu của các enzym giới hạn và các gen chỉ thị dễ nhận biết. Các plasmid tự nhiên

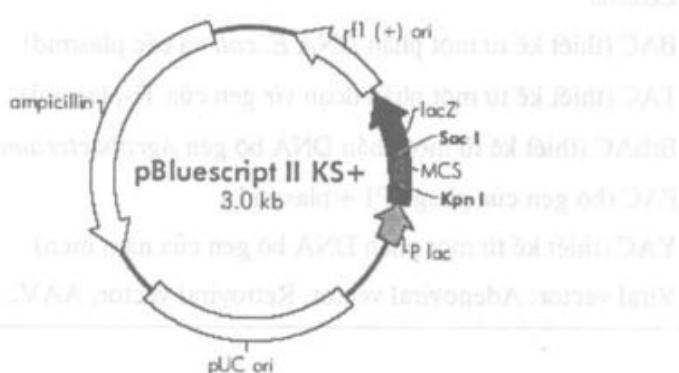
tách chiết được từ vi khuẩn như pSC101 (Stanley và Cohen, 1973), ColE1 (Hershfield và cộng sự 1974)...

Nhiều loại plasmid nhân tạo (thể hệ 2 và thể hệ 3) được tạo nên bằng cách tập hợp các đặc tính quý của nhiều plasmid tự nhiên, gắn thêm các gen chỉ thị và đoạn đa cắt nối (MCS - many cloning site) tạo nên các plasmid mạnh.



Hình 4. 4. Vector tách dòng pBR 22 và pUC18

Plasmid thế hệ 2 điển hình là pBR322, pUC18, pSP64, pSP56, Gemini...  
Plasmid thế hệ 3 điển hình là pCR2.1, pBluescript, pET...



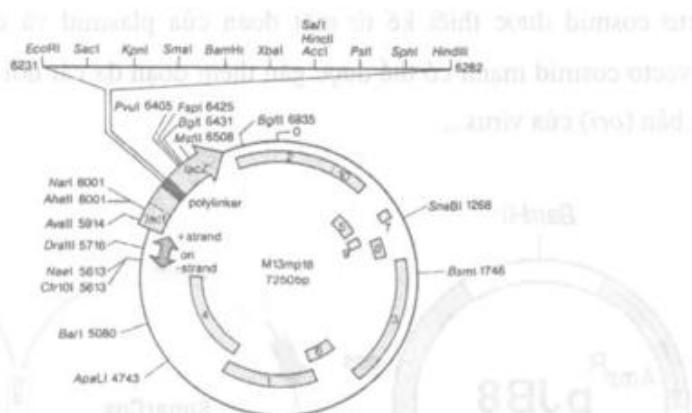
Hình 4. 5. Cấu trúc vector pBluescript II KS +

**b. Vector tách dòng là phage**

Tóm tắt về Vector tách dòng là phage

Phage (Bacteriophage) bao gồm nhiều loại thực khuẩn thể có bộ gen DNA mạch đơn hoặc mạch kép được sử dụng làm vector tách dòng gồm nhiều loại khác nhau: f1, M13, fd, λ EMBL3, EMBL4, λGEM 11, λGEM12, λgt11, λ Lgt10, λ LZAP... Các vector tách dòng là phage thường được tạo nên từ sự cải biến bộ gen phage. Ví dụ, phage M13 gắn thêm gen lacZ, và các vị trí cắt đặc hiệu của các loại enzym giới hạn tạo thành nhiều loại vector tách dòng khác nhau: phage M13mp1, M13mp7, M13mp18...

Sử dụng vector tách dòng phage có một số ưu điểm hơn so với plasmid như vector phage cho cài gắn đoạn DNA kích thước lớn hơn plasmid, phage có hệ thống gen giúp xâm nhiễm và tạo nên số bản sao rất cao trong tế bào vi khuẩn. Vector tách dòng là phage được sử dụng làm vector tách dòng hiệu quả ở cả tế bào chủ là *prokaryote* và các tế bào chủ là *eukaryote*.



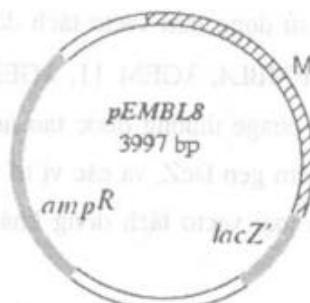
Hình 4. 6. Cấu trúc vector phage M13mp18

**c. Vector tách dòng là phagemid**

Phagemid cấu tạo gồm một số gen của plasmid (các gen chỉ thị, đoạn ori...)

và một phần gen của phage (đoạn ori, các gen cần thiết cho sự đóng gói).

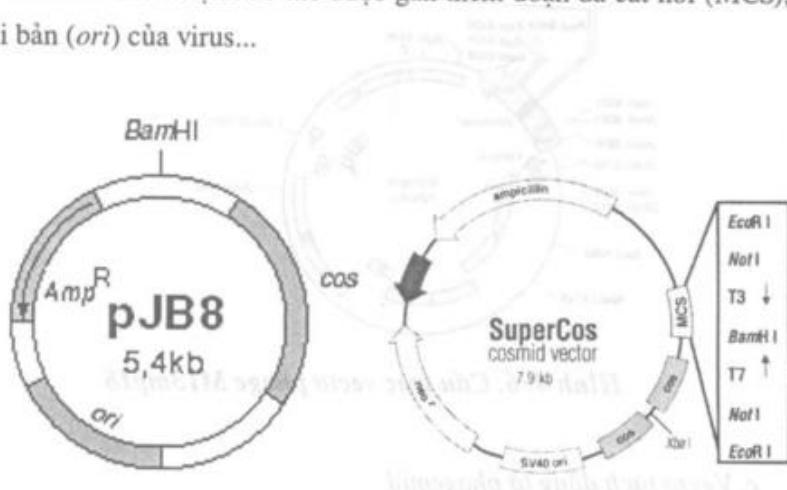
Ví dụ, vectơ pEMBL8 là phagemid cấu tạo từ một đoạn của phage M13, mang các đoạn gen chỉ thị kháng sinh *Amp<sup>R</sup>* và *lacZ* lấy từ plasmid.



Hình 4. 7. Vectơ phagemid pEMBL8

#### d. Vectơ tách dòng là cosmid

Vectơ cosmid được thiết kế từ một đoạn của plasmid và đoạn *cos* của phage. Các vectơ cosmid mạnh có thể được gắn thêm đoạn đa cắt nối (MCS), đoạn khởi đầu tái bản (*ori*) của virus...



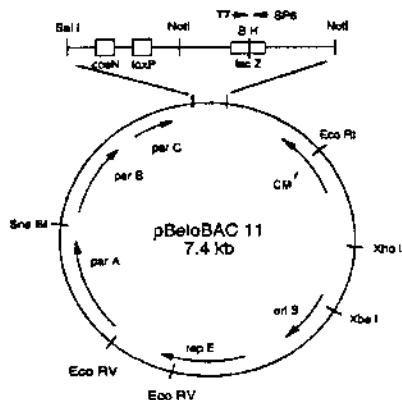
Hình 4. 8. Cấu trúc vectơ cosmid pJB8 và SuperCos

Ví dụ, vectơ SuperCos kích thước 7,9 kb có cấu trúc gồm 2 đoạn cos của bộ gen phage, 2 gen chỉ thị kháng sinh (ampicillin và neomycin), đoạn *ori* của virus SV40 và đoạn đa cắt nối (MCS). Vectơ cosmid cho cài gắn DNA insert có kích thước lớn tới 50 kb.

#### e. Vectơ tách dòng là các nhiễm sắc thể nhân tạo

Bộ gen của sinh vật bậc cao có kích thước rất lớn. Do đó, khi thiết lập các ngân hàng bộ gen, ngân hàng cDNA sử dụng các loại vectơ plasmid, phage, phagemid, cosmid... tạo nên số lượng dòng rất lớn, gây khó khăn trong tách dòng. Các nhiễm sắc thể nhân tạo được thiết kế làm vectơ tách dòng, cho phép cài gắn các đoạn DNA insert có kích thước rất lớn từ 100 kb đến hàng nghìn kb, tạo thuận lợi trong tách dòng và lập ngân hàng bộ gen. Mỗi loại vectơ tách dòng là nhiễm sắc thể nhân tạo có những đặc điểm đặc trưng riêng, cho phép cài gắn đoạn DNA có kích thước khác nhau.

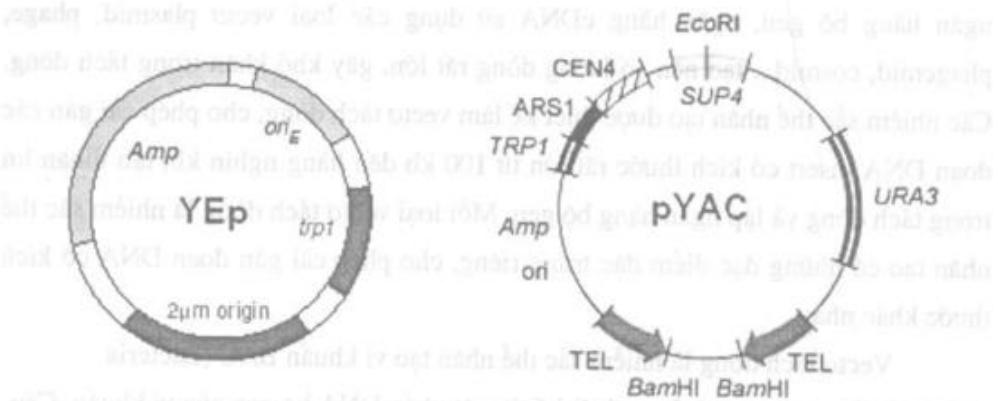
Vectơ tách dòng là nhiễm sắc thể nhân tạo vi khuẩn BAC (Bacteria Artificial Chromosomes) được thiết kế từ một phần DNA bộ gen của vi khuẩn. Cấu trúc vectơ BAC gồm các đoạn *ori* (origin replication), các gen chỉ thị đặc hiệu, đoạn đa cắt nối (MCS - many cloning site) và promoter đặc hiệu.



**Hình 4. 9. Vectơ nhiễm sắc thể nhân tạo vi khuẩn pBeloBAC11**

**Vector BAC** cho phép cài gán đoạn DNA kích thước từ 100 kb - 300 kb. Vector BAC được ứng dụng nhiều trong lập ngân hàng bộ gen và giải trình tự bộ gen của sinh vật *prokaryote* và *eukaryote*.

Vector tách dòng là nhiễm sắc thể nhân tạo nấm men (Yeast Artificial Chromosomes -YAC) có thể cho cài gán đoạn DNA lớn tới 2 000 kb. Vector YAC được thiết kế từ một số trình tự đặc hiệu của bộ gen nấm men, trong đó có 3 trình tự quan trọng là trình tự tái bản ARS (autonomous replicating sequences), tâm động CEN (centromer) và trình tự TEL (telomer).



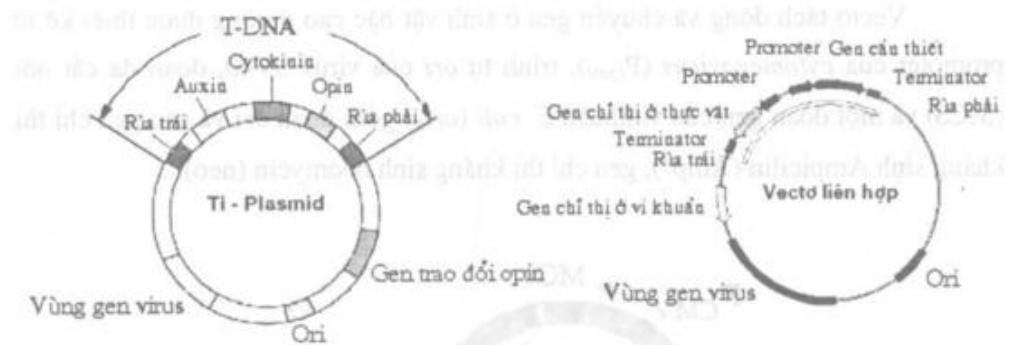
Hình 4.10. Vector nhiễm sắc thể nhân tạo nấm men YEplac và pYAC

Ngoài các trình tự chủ yếu ARS, CEN, TEL... vector YAC còn được gắn thêm các gen chỉ thị đặc hiệu với tế bào nấm men *trpI* (auxotrophs), *ura 3* (uracil), *leu 2* (leucin), *his 3* (histidin)... các gen chỉ thị kháng sinh và các đoạn *ori* của vi khuẩn hoặc đoạn đa cắt nối (MCS)...

Loại vector YAC đơn giản nhất là YEplac (yeast episomal plasmid). Vector YEplac mang đoạn *ori* và phần lớn các gen trong plasmid *2μm* của tế bào nấm men, và được gắn thêm đoạn *ori* của *E. coli*, các gen chỉ thị kháng sinh (*Amp<sup>R</sup>*), gen chỉ thị đặc hiệu với tế bào nấm men *trp I* (auxotrophs). Vector YEplac cho phép tạo số bản sao cao từ 10 - 40 trong một tế bào.

Vector pYAC (Yeast Artificial Centromere) gồm các trình tự ARS, CEN, TEL, các đoạn *ori* và các gen chỉ thị đặc hiệu. Vector pYAC có trình tự tái bản tự động ARS (Autonomously Replicating Sequences) thường tạo số bản sao thấp khoảng 2 - 4 bản sao trong một tế bào, nhưng rất bền vững trong tế bào chủ.

*Ti plasmid* là plasmid có kích thước lớn tới 200 kb trong tế bào vi khuẩn *Agrobacterium tumefaciens*. Ti plasmid có một đoạn T - DNA kích thước khoảng 10 - 20 kb mang các gen tổng hợp opin, auxin... gây nên các khối u ở thực vật hai lá mầm. Do khả năng xâm nhiễm tự nhiên vào cây hai lá mầm thông qua các vết xước rẽ cây, nên Ti plasmid được cải biến cấu trúc làm vector tách dòng và vector chuyển gen có hiệu quả ở thực vật.



Hình 4.11. Sơ đồ Ti - Plasmid và vector liên hợp

Ti plasmid kích thước quá lớn, gây khó khăn trong tách dòng gen, do đó người ta sử dụng một phần DNA của Ti plasmid, gắn thêm các gen chỉ thị và các đoạn DNA chức năng tạo nên các loại vector tách dòng vector nhị thể, vector liên hợp. Ví dụ, pTiC58 (kích thước khoảng 194 000 bp, Depicker và cộng sự 1980), pTiAch5

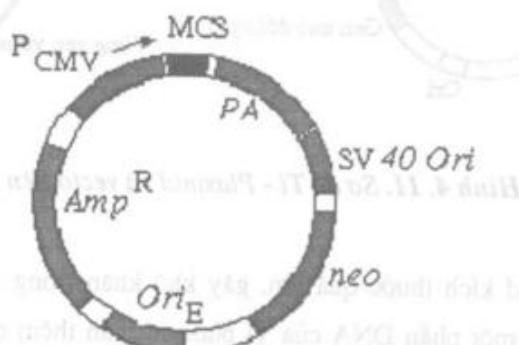
(kích thước khoảng 213 000 bp, Devos và cộng sự 1981) được sử dụng rộng rãi trong tách dòng và chuyển gen ở thực vật có hiệu quả cao.

#### *h. Vector tách dòng ở tế bào sinh vật eukaryote*

Nhiều loại vectơ tách dòng plasmid, cosmid... hoạt động kém ở các tế bào thực vật và động vật bậc cao. Do vậy, trong tách dòng gen và chuyển gen ở động vật có vú và nhiều loài sinh vật bậc cao khác, người ta thường sử dụng vectơ tách dòng là các loại virus được cải biến bộ gen, cắt bỏ các gen độc lực và cài gắn thêm các gen chỉ thị đặc hiệu.

Các loại virus SV40 (*Simian virus*), *adenovirus*, *retrovirus*, *baculovirus* và *virus herpes*... cải biến bộ gen (loại bỏ các gen độc, thêm các gen chỉ thị và các đoạn MCS...) được sử dụng làm vectơ tách dòng, vectơ chuyển gen ở sinh vật bậc cao.

Vectơ tách dòng và chuyển gen ở sinh vật bậc cao thường được thiết kế từ promoter của *cytomegavirus* ( $P_{CMV}$ ), trình tự *ori* của virus SV40, đoạn đa cắt nối (MCS) và một đoạn gen của vi khuẩn *E. coli* ( $ori_E$ ) gồm đoạn *ori* và các gen chỉ thị kháng sinh Ampicillin ( $Amp^R$ ), gen chỉ thị kháng sinh neomycin ( $neo$ )...



Hình 4.12. Sơ đồ vectơ tách dòng ở sinh vật bậc cao

### 5. Vector biểu hiện gen

#### 5. 1. Đặc điểm cấu trúc vector biểu hiện gen

Vector biểu hiện gen (expression vector) là những vector tách dòng mang các promoter mạnh, cho phép biểu hiện đồng thời cả gen chỉ thị và gen tách dòng tạo nên các protein lai (fusion protein).

Một vector biểu hiện gen gồm các thành phần chủ yếu: promoter mạnh, trình tự sao chép (*ori*), vị trí khởi đầu phiên mã, vị trí bám của riboxom, tín hiệu kết thúc dịch mã, các trình tự đa cắt nối (MCS) và các gen chỉ thị...



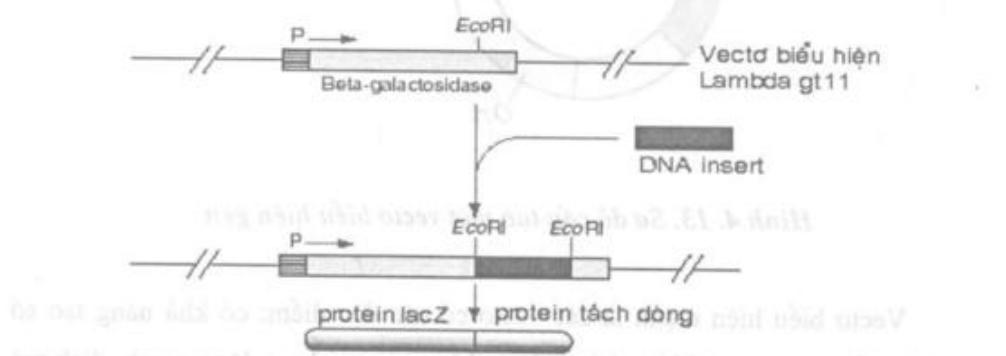
Hình 4. 13. Sơ đồ cấu tạo một vector biểu hiện gen

Vector biểu hiện mạnh là các vector có các đặc điểm: có khả năng tạo số lượng lớn bản sao trong tế bào chủ (*ori* mạnh); promoter hoạt động mạnh, dịch mã tạo nhiều protein lai; đoạn trình tự MCS có nhiều vị trí cắt đặc hiệu của nhiều loại enzym giới hạn; vector vẫn giữ được hoạt tính khi cài gắn gen tách dòng (đoạn DNA insert) kích thước lớn; hoạt động của vector không ảnh hưởng hoặc gây ức chế tế bào chủ (host cell); gen chỉ thị dễ dàng nhận biết...

### 5. 2. Các loại promoter thường sử dụng trong vectơ biểu hiện gen

Các loại vectơ biểu hiện gen trong tế bào *prokaryote* thường sử dụng các promoter mạnh của thực khuẩn thể (T7 promoter, T3 promoter,...) hoặc một số promoter mạnh của vi khuẩn (*lac* promoter, *Trp* promoter, *lac* UV5 promoter...) là những loại promoter thích hợp với nhiều loại RNA polymerase. Ví dụ, T7 promoter kiểm soát sự biểu hiện các gen muộn ở bacteriophage T7, được sử dụng nhiều trong thiết kế các vectơ biểu hiện gen. Vectơ biểu hiện gen có T7 promoter cho phép biểu hiện gen tách dòng ở các tế bào chủ là vi khuẩn (*E. coli*, *bacillus*...) khi có mặt các chất cảm ứng (IPTG, lactose...).

Biểu hiện trong tế bào *eukaryote* thường sử dụng các vectơ biểu hiện gen có các promoter mạnh của virus CMV promoter (cytomegavirus), SV40 promoter (Simian virus)... hoặc các promoter mạnh gen nấm men như PGK promoter (gen photphoglyxerate), ADH I promoter (gen alcoholdehydrogenase I), GAL promoter (gen tổng hợp galactose), AOX I promoter (gen alcohol oxydase)...



**Hình 4. 14. Sơ đồ vị trí gắn gen tách dòng vào vectơ biểu hiện lambda gt11 và sự hình thành protein lai**

Trong vectơ biểu hiện gen, vị trí cài gắn đoạn DNA tách dòng có thể ở

ngay sau promoter hoặc ở trước vị trí kết thúc phiên mã của gen chỉ thị. Tuỳ theo

cấu trúc đặc trưng của đoạn gen tách dòng và loại promoter để lựa chọn vị trí gắn gen tách dòng thích hợp. Ví dụ, sử dụng vectơ tách dòng có gen lacZ, thì gen cần tách dòng thường được gắn ở cuối gen lacZ.

### 6. Các hệ thống biểu hiện gen (Expression systems)

Biểu hiện gen tạo sản phẩm protein tái tổ hợp là kỹ thuật quan trọng, quyết định sự thành bại của kỹ thuật gen. Hệ thống các tế bào chủ (host cell) thường được sử dụng trong biểu hiện gen gồm: tế bào vi khuẩn, tế bào nấm men, tế bào trứng động vật có vú và *Baculovirus*.

#### 6. 1. Biểu hiện gen trong tế bào vi khuẩn

Vì khuẩn *E. coli* là tế bào chủ được ưa chuộng nhất trong biểu hiện gen, do mang nhiều ưu điểm hơn so với các tế bào khác:

- Vì khuẩn *E. coli* có tốc độ sinh trưởng nhanh, khả năng biểu hiện protein tái tổ hợp mạnh, chỉ sau 8 giờ nuôi cấy ở điều kiện thích hợp đã có sản phẩm protein tái tổ hợp. Khả năng tạo sản phẩm gen cao từ 50 mg / lít đến 500 mg/lít,
- Biểu hiện gen trong tế bào *E. coli* có thể giảm được các chi phí cho công nghệ và hóa chất, nên giá thành sản phẩm hạ.

- Cấu trúc bộ gen *E. coli* và các đặc điểm di truyền đã được biết tương đối đầy đủ, tạo nên nhiều thuận lợi khi biểu hiện protein tái tổ hợp ở *E. coli*.

Có rất nhiều chủng vi khuẩn *E. coli* được sử dụng trong biểu hiện gen, trong đó chủng *E. coli* BL 21 là một trong những chủng *E. coli* được ưa chuộng và sử dụng rộng rãi để biểu hiện nhiều loại gen khác nhau.

Tuy nhiên, sử dụng vi khuẩn *E. coli* làm tế bào chủ trong biểu hiện gen có một số hạn chế sau:

- Các gen mã hoá các loại protein có kích thước lớn hơn 50 kD, protein giàu cystein (ví dụ như gen mã hoá plasminogen) hoặc những sản phẩm protein có

sự hình thành nhiều liên kết disulfua (S - S) rất khó khi biểu hiện gen trong tế bào *E. coli*.

- Tế bào *E. coli* có thể cho biểu hiện tốt các loại protein không biến đổi cấu trúc phân tử sau dịch mã và không có quá trình glycosyl hoá.
- Một số chủng vi khuẩn *E. coli* có thể gây bệnh, hoặc tạo các độc tố khi biểu hiện gen của sinh vật bậc cao.
- Một khác, khả năng tạo và tiết protein ngoại bào của các chủng vi khuẩn *E. coli* tương đối thấp.

Trong kỹ thuật gen, vi khuẩn *Bacillus* cũng được sử dụng làm tế bào biểu hiện gen, vi khuẩn *Bacillus* có ưu điểm là có khả năng tạo và tiết protein ngoại bào mạnh, khả năng tạo sản phẩm tương đối cao, chi phí cho công nghệ thấp. Vi khuẩn *Bacillus* được sử dụng trong biểu hiện gen chủ yếu là *B. subtilis*, tuy nhiên mức độ ưa chuộng kém hơn vi khuẩn *E. coli*.

## **6. 2. Biểu hiện gen trong *Baculovirus* nuôi trong tế bào côn trùng**

*Baculovirus* là nhóm virus ký sinh hơn 600 loài côn trùng khác nhau, *baculovirus* là loại tế bào chủ cho phép biểu hiện các gen mã hoá các loại protein có kích thước lớn hơn 50 kD. *Baculovirus* cài biến gen thường được cài gắn thêm các promoter cực mạnh (super strong promoter) của gen mã hoá protein vỏ virus. *Baculovirus* nuôi cấy trong tế bào trùng, cho phép biểu hiện nhiều loại protein người như  $\alpha$  - interferon,  $\beta$  - interferon, adenosin deaminase, erythropoietin, interleukin 2, plasminogen activator (TPA)...

Biểu hiện gen trong *Baculovirus* có nhiều đặc điểm thuận lợi: *Baculovirus* cho phép biểu hiện các protein có phản ứng glycosyl hoá chính xác và có phản ứng cắt chuỗi peptid tín hiệu. Sử dụng *Baculovirus* trong biểu hiện gen có hiệu quả cao, giá thành sản phẩm protein tái tổ hợp giảm.

Hạn chế lớn nhất khi biểu hiện gen ở *Baculovirus* là sự tạo thành protein tái tổ hợp chậm, thường sau 4 - 5 ngày nuôi cấy mới có sản phẩm, thời gian nuôi cấy

lâu (khoảng 10-15 ngày). Mặt khác, điều kiện thực hiện công nghệ và môi trường nuôi cấy phức tạp.

### 6. 3. Biểu hiện gen trong tế bào nấm men

Nấm men là nhóm tế bào chủ được sử dụng rộng rãi trong biểu hiện gen sinh vật bậc cao. Nấm men có nhiều ưu điểm: cho phép sử dụng cả hai loại vectơ biểu hiện trong *prokaryote* và trong *eukaryote*, biểu hiện protein kích thước lớn hơn 50 kD, cấu trúc bộ gen của nấm men đã được nghiên cứu đầy đủ, không gây bệnh. Nấm men có khả năng tạo sản phẩm nhanh, chỉ sau 8 giờ nuôi cấy đã có sản phẩm protein tái tổ hợp, hiệu xuất biểu hiện gen tương đối cao (từ 50 mg đến 5 000 mg/lít dịch nuôi cấy).

Các vectơ biểu hiện gen trong tế bào nấm men thường có promoter AOX (gen alcohol oxidase) rất mạnh, có khả năng kiểm soát gen cần biểu hiện cao. Biểu hiện gen trong tế bào nấm men có thể thực hiện các phản ứng glycosyl hoá chính xác, phản ứng cắt chuỗi tín hiệu hoặc thay đổi cấu trúc protein sau dịch mã. Mặt khác, nhiều loại protein giàu cystein, hoặc có nhiều liên kết disulfid cũng được biểu hiện trong tế bào nấm men.

Nấm men *S. Cceseivisiae* và *Pichia Pastoris* là các tế bào biểu hiện gen được ưa chuộng nhất trong các hệ thống biểu hiện gen.

### 6. 4. Biểu hiện gen trong tế bào động vật bậc cao

Trong sản xuất vaccine tái tổ hợp và nhiều loại protein có hoạt tính sinh học như interferon γ, hGH, TPA, Interleukin 2, factor VIII, factor IX,... người ta thường sử dụng hệ thống biểu hiện gen là các tế bào S2 của ruồi giấm và tế bào trứng chuột đất vàng Trung Quốc (CHO - Chinese Hamster Ovary).

Tế bào động vật bậc cao cho phép biểu hiện các protein kích thước lớn hơn 50 kD. Các vectơ biểu hiện gen thường có các promoter mạnh SV 40, CMV cho phép thực hiện các phản ứng glycosyl hoá, hoặc cắt chuỗi tín hiệu chính xác. Tế

bào trứng chuột đất vàng Trung Quốc được ứng dụng trong sản xuất nhiều loại protein tái tổ hợp sử dụng cho con người, do khả năng tạo thành các protein có cấu trúc không gian thích hợp...

Sử dụng hệ thống biểu hiện gen là tế bào động vật bậc cao có một số điểm hạn chế: chỉ có thể biến nạp bằng xung điện, thời gian sàng lọc rất lâu, nuôi cấy phức tạp, giá thành cao.

Mỗi loại hệ thống biểu hiện gen có đặc điểm đặc trưng riêng, cùng với các thuận lợi và hạn chế nhất định. Cần căn cứ vào đặc điểm của gen biểu hiện, bản chất của protein tách dòng và protein lai (fusion protein), cũng như điều kiện cụ thể của từng phòng thí nghiệm để lựa chọn hệ thống biểu hiện gen thích hợp. Thông thường để biểu hiện protein có kích thước dưới 30 kD, thường chọn hệ thống biểu hiện trong tế bào vi khuẩn. Các protein có kích thước lớn, có quá trình biến đổi sau dịch mã, hoặc glycosyl hoá có thể chọn hệ thống biểu hiện là nấm men, baculovirus hoặc tế bào CHO...

## **II. Nguyên lí kỹ thuật gen**

### **1. Các bước chủ yếu của kỹ thuật gen**

Kỹ thuật gen gồm nhiều giai đoạn khác nhau nhằm biến đổi cấu trúc gen tạo gen mới, đưa các gen tái tổ hợp vào tế bào chủ (host cell) để thu nhận các protein tái tổ hợp, hoặc nghiên cứu cấu trúc bộ gen và lập bản đồ trình tự gen của các loài sinh vật. Kỹ thuật gen có thể chia làm các bước cơ bản sau:

#### **+ Chuẩn bị gen cần tách dòng:**

Gen cần tách dòng còn gọi là gen cần thiết hay đoạn DNA insert, có thể được chuẩn bị bằng phương pháp tách chiết (DNA, RNA) từ các sinh vật mang nguồn gen tự nhiên hoặc tổng hợp nhân tạo.

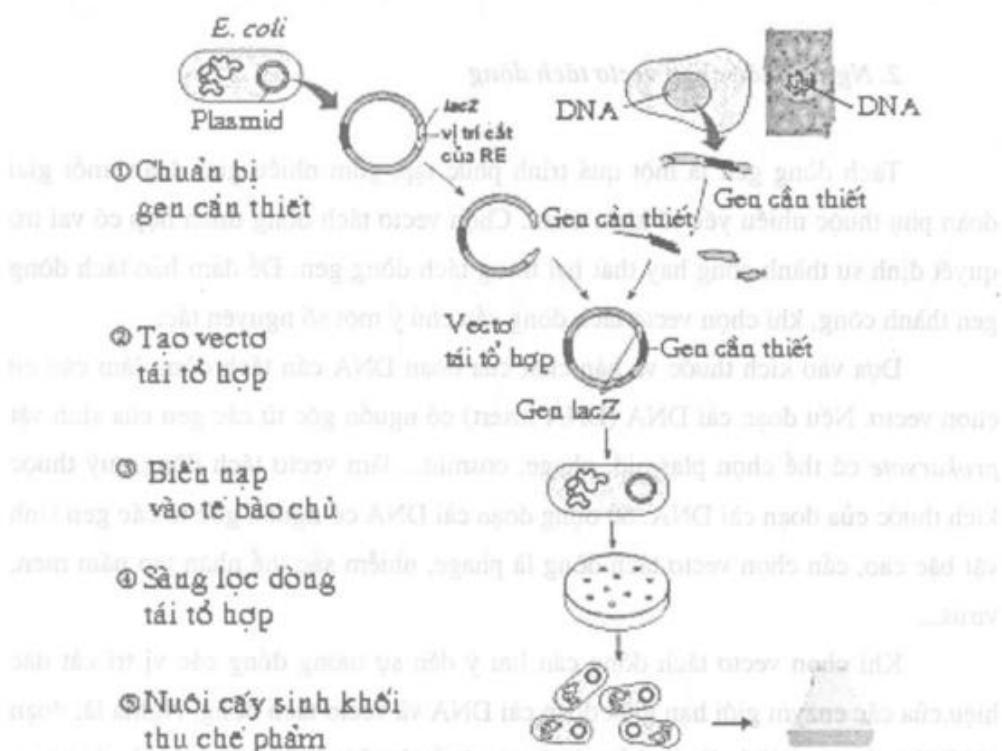
Tách chiết (DNA, RNA) từ các sinh vật mang nguồn gen tự nhiên: Thực hiện kỹ thuật tách chiết DNA, RNA của các tế bào, mô sinh vật bậc cao hoặc tách chiết plasmid từ vi khuẩn. Tinh sạch và chọn lọc để thu được các gen cần tách

dòng. Sử dụng kỹ thuật nhân gen PCR để nhân gen cần thiết (gen cần tách dòng) lên một số lượng lớn các bản sao, đảm bảo một lượng DNA (RNA cần thiết).

Có thể tổng hợp nhân tạo đoạn DNA insert trên cơ sở những trình tự đặc hiệu đã biết, hoặc từ trình tự các acid amin trong chuỗi peptid đã biết để dự đoán trình tự đoạn gen. Thực hiện kỹ thuật tổng hợp nhân tạo đoạn oligonucleotid mạch đơn và kỹ thuật nhân gen PCR để tạo số lượng các đoạn DNA insert (gen cần tách dòng) đủ lớn.

#### + Tao vectơ tái tổ hợp:

Dựa vào đặc điểm của gen cần tách dòng để lựa chọn loại vectơ tách dòng và các enzym giới hạn phù hợp, nhằm thực hiện phản ứng ghép nối giữa các gen cần thiết với vectơ tách dòng tạo vectơ tái tổ hợp.



Hình 4. 15. Sơ đồ các bước chủ yếu của kỹ thuật gen

+ Biến nạp vectơ tái tổ hợp vào tế bào chủ:

Sử dụng các phương pháp biến nạp khác nhau (sốc nhiệt, xung điện, bắn gen...) để đưa các vectơ tái tổ hợp vào tế bào chủ (host cell). Tạo các điều kiện môi trường thích hợp để vectơ tái tổ hợp tồn tại và được nhân lên cùng tế bào chủ.

+ Sàng lọc các dòng tế bào mang vectơ tái tổ hợp:

Sàng lọc dòng tái tổ hợp (screening) là quá trình kiểm tra, chọn lọc và tách riêng các dòng tái tổ hợp (dòng tế bào chủ mang vectơ tái tổ hợp) có hoạt tính.

+ Nuôi cấy tế bào tái tổ hợp để thu sản phẩm:

Tế bào tái tổ hợp được nuôi trong các bình lén men, với điều kiện thích hợp về dinh dưỡng, nhiệt độ, độ pH,... để các tế bào tăng sinh khối nhanh, tạo các sản phẩm protein tái tổ hợp ở mức cao nhất. Thực hiện công nghệ tách chiết protein tái tổ hợp, tinh sạch thu chế phẩm.

## 2. Nguyên tắc chọn vectơ tách dòng

Tách dòng gen là một quá trình phức tạp, gồm nhiều giai đoạn, mỗi giai đoạn phụ thuộc nhiều yếu tố khác nhau. Chọn vectơ tách dòng thích hợp có vai trò quyết định sự thành công hay thất bại trong tách dòng gen. Để đảm bảo tách dòng gen thành công, khi chọn vectơ tách dòng cần chú ý một số nguyên tắc:

Dựa vào kích thước và bản chất của đoạn DNA cần tách dòng làm căn cứ chọn vectơ. Nếu đoạn cài DNA (DNA insert) có nguồn gốc từ các gen của sinh vật *prokaryote* có thể chọn plasmid, phage, cosmid... làm vectơ tách dòng, tùy thuộc kích thước của đoạn cài DNA. Sử dụng đoạn cài DNA có nguồn gốc từ các gen sinh vật bậc cao, cần chọn vectơ tách dòng là phage, nhiễm sắc thể nhân tạo nấm men, virus...

Khi chọn vectơ tách dòng cần lưu ý đến sự tương đồng các vị trí cắt đặc hiệu của các enzym giới hạn giữa đoạn cài DNA và vectơ tách dòng. Nghĩa là, đoạn cài DNA và vectơ tách dòng phải có các vị trí cắt đặc hiệu của cùng một loại enzym giới hạn (để đảm bảo quá trình cắt và nối đoạn cài DNA vào vectơ tách dòng dễ dàng và thuận lợi).

Vectơ tách dòng phải thích hợp với loại tế bào chủ, có khả năng tái bản trong tế bào chủ, tạo số lượng bản sao lớn (lớn hơn 50 bản sao trong một tế bào), nhưng không gây ảnh hưởng đến hoạt động của tế bào chủ.

Gen chỉ thị trong vectơ tách dòng càng dễ nhận biết càng tốt (chỉ thị màu hoặc chỉ thị kháng sinh) để dễ dàng chọn lọc các vectơ tái tổ hợp. Các điểm cắt của enzym giới hạn hoặc vùng đa cắt nối MCS (many cloning site) thường nằm trong các gen chỉ thị, giúp cho sự nhận biết và sàng lọc thể tái tổ hợp thuận lợi.

Vectơ tách dòng phải có hiệu quả tách dòng (cloning efficiency) cao, nghĩa là tạo nên tỉ lệ vector tái tổ hợp cao.

### *3. Nguyên tắc tạo vectơ tái tổ hợp*

Tạo vectơ tái tổ hợp là thực hiện kỹ thuật gắn nối gen cần tách dòng vào vectơ tách dòng. Tạo vectơ tái tổ hợp có thể thực hiện theo các bước chủ yếu sau:

Chuẩn bị nguyên liệu: Gen cần tách dòng (DNA insert) tinh sạch cần được nhân bằng PCR tạo một hàm lượng cần thiết sản phẩm PCR (thông thường thực hiện PCR với khoảng 20 - 25 chu kỳ). Vectơ tách dòng thích hợp (ví dụ, pBR 322, pBluescript, pUC18...) có hàm lượng đủ lớn, được mở vòng bằng enzym giới hạn phù hợp.

Thực hiện phản ứng nối gen cần tách dòng với vectơ: chuẩn bị thành phần phản ứng phù hợp với mỗi loại vectơ tách dòng. Ví dụ, vectơ tách dòng là pBluescript hoặc pUC 18 có thể chuẩn bị hỗn hợp phản ứng gồm các thành phần: 1,0 µl đệm phản ứng; 0,5 µl ATP 10 mM; 1,0 µl enzym T4 DNA ligase ; 1,0 µl vectơ pBluescript (pUC18) ; 5,0 µl sản phẩm PCR gen cần tách dòng ; 1,5 µl H<sub>2</sub>O.

Hỗn hợp các thành phần phản ứng được ủ qua đêm ở nhiệt độ 16°C. Sau đó, tiến hành điện di với thang DNA chuẩn để kiểm tra kết quả tạo vectơ tái tổ hợp. Vectơ tái tổ hợp đã tạo được cần giữ trong tủ lạnh sâu ở - 80°C đến - 85°C để tiếp tục các nghiên cứu tiếp theo.

#### **4. Một số kỹ thuật biến nạp vectơ tái tổ hợp vào tế bào chủ**

Biến nạp vectơ tái tổ hợp vào tế bào chủ được thực hiện bằng nhiều kỹ thuật khác nhau: bắn gen, vi tiêm, xung điện, sốc nhiệt, sử dụng PEG...

Hiệu quả biến nạp phụ thuộc vào loại tế bào chủ, kích thước vectơ tái tổ hợp và kỹ thuật biến nạp thích hợp. Để hiệu quả biến nạp cao, cần lựa chọn kỹ thuật biến nạp thích hợp với mỗi loại tế bào chủ. Ví dụ, khi sử dụng tế bào chủ là vi khuẩn thường sử dụng phương pháp biến nạp bằng kỹ thuật sốc nhiệt, sử dụng tế bào chủ là nấm men thì kỹ thuật xung điện là phương pháp biến nạp có hiệu quả cao, dễ thực hiện...

##### **4. 1. Biến nạp vectơ tái tổ hợp vào tế bào *E. coli* bằng kỹ thuật sốc nhiệt**

Tế bào chủ là vi khuẩn *E. coli* bao gồm nhiều chủng khác nhau (DH5 $\alpha$ , LE 392, MC 1061, BL 21...) thường được sử dụng trong các phòng thí nghiệm công nghệ sinh học phân tử. Kỹ thuật biến nạp vào tế bào vi khuẩn với CaCl<sub>2</sub> và sốc nhiệt được Mandel và Hinga phát hiện năm 1970, sau đó được hoàn thiện bởi nhiều tác giả ở các phòng thí nghiệm khác nhau. Để biến nạp các plasmid vào tế bào *E. coli* bằng phương pháp sốc nhiệt, cần chuẩn bị tế bào khả biến trước khi thực hiện kỹ thuật biến nạp.

###### **a. Chuẩn bị tế bào khả biến**

Tế bào *E. coli* (DH5 $\alpha$ , LE 392, MC 1061, BL 21...) được nuôi cấy trên môi trường thạch LB (5% cao nấm men, 10% trypton, 10% NaCl). Lấy 1 khuỷn lắc *E. coli* vào mỗi 5 ml môi trường LB lòng, nuôi ở 37°C, lắc với tốc độ 200 vòng / phút trong 12 - 16 giờ. Lấy 1 ml dịch VK + 99 ml môi trường LB nuôi lắc 2 - 3 giờ, khi OD<sub>600</sub> đạt 0,4 - 0,6 lấy dịch nuôi cấy để lên đá trong 1,0 - 1,5 giờ, sau đó li tâm lạnh thu cạn tế bào ở tốc độ 4 500 - 5 000 vòng trong 8 - 10 phút.

Hoà cặn tế bào trong 100 ml CaCl<sub>2</sub> 100 mM lạnh, để trên khay đá 15 phút rồi li tâm lạnh thu cặn tế bào. Cặn tế bào hòa trong 40 ml CaCl<sub>2</sub> 100 mM lạnh, để 1 giờ trên đá, sau đó li tâm thu cặn tế bào với tốc độ 4 000 vòng trong 10 phút, thêm 500 µl CaCl<sub>2</sub> 100 mM lạnh, có chứa 15% glycerol vào cặn tế bào thu được các tế bào khả biến. Lấy vào mỗi ống eppendorf 50 µl tế bào khả biến, giữ trong tủ lạnh sâu ở nhiệt độ -75°C đến -80°C để thực hiện biến nạp.

b. Kỹ thuật biến nạp

Lấy 50µ tế bào khả biến đang giữ trong ống eppendorf ở nhiệt độ -75°C đến -80°C (trong tủ lạnh sâu) để trên đá khoảng 15 phút, sau đó thêm 0,5 ng DNA (vectơ tái tổ hợp) đảo nhẹ ống 3 - 4 lần và để trên đá khoảng 15 - 20 phút.

Lấy mẫu đặt vào bể ấm nhiệt có nhiệt độ 42°C trong 2 phút, đặt nhanh lên đá 2 phút. Sau đó thêm khoảng 1 000 µl môi trường LB lạnh, nuôi trong 37°C ở máy lắc trong 30 - 60 phút. Lấy 100 µl dịch tế bào đã súc nhiệt, trải lên hộp Petri môi trường LB 2% agarose có bổ sung ampicilin với nồng độ 50 µg / ml , IPTG và X- gal. Ủ hộp Petri ở 37°C trong 12 - 16 giờ, kiểm tra các khuẩn lạc phát triển trên hộp Petri. Thu nhận khuẩn trắng, cấy vào môi trường lỏng để kiểm tra kết quả.

Biến nạp *E. coli* bằng kỹ thuật súc nhiệt có hiệu quả khoảng 1/ 10 000. Sàng lọc các tế bào mang vectơ tái tổ hợp, chuyển sang ống thạch nghiêng để giữ chủng giống.

4. 2. Biến nạp vectơ tái tổ hợp vào tế bào nấm men bằng kỹ thuật xung điện

a. Chuẩn bị tế bào

Cấy 1 khuẩn lạc nấm men vào 5 ml môi trường YPD (1% cao nấm men, 2% pepton, 2% glucose), nuôi ở 28°C với tốc độ lắc 250 vòng/ phút trong 1 ngày đêm. Lấy 30 µl dịch nuôi cấy vào 100ml YPD, lắc 16 - 18 giờ, đo OD<sub>600</sub> để xác

định mật độ tế bào ( $OD = 1,3 - 1,5$ ). Để mẫu trên đá 5 - 10 phút, li tâm với tốc độ 4 500 - 5 000 vòng trong 5 phút, lấy cặn tế bào hoà trong 100 ml nước khử ion lạnh, li tâm thu cặn tế bào. Hoà tế bào trong 20 ml sorbitol 1M, li tâm 4 500 vòng trong 5 phút thu cặn tế bào. Hoà cặn tế bào với 0,2 ml sorbitol, chia đều mỗi ống eppendorf 50  $\mu$ l dịch tế bào khả biến giữ trong tủ lạnh sâu -75°C đến -80°C.

b. Kỹ thuật xung điện

Lấy 50  $\mu$ l dịch tế bào khả biến trong ống eppendorf giữ trong tủ lạnh sâu thêm vào 10 -15  $\mu$ g DNA plasmid tái tổ hợp, trộn đều và chuyển vào cuvet xung điện 0,2 cm. Thực hiện xung điện 0,25  $\mu$ F, 200  $\Omega$ , 1,5 kV trong 4 - 5 phần nghìn giây. Bổ sung 1 ml sorbitol 1M lạnh, đảo đều chuyển sang ống mới, cấy lên môi trường thạch đĩa để kiểm tra kết quả biến nạp. Hiệu quả biến nạp bằng xung điện tương đối cao (khoảng 1%).

5. *Kỹ thuật chọn lọc các dòng tế bào tái tổ hợp*

Dòng tế bào tái tổ hợp (còn gọi là thể tái tổ hợp) là dòng các tế bào mang vectơ tái tổ hợp, được nuôi ở môi trường thích hợp tạo thành các khuẩn lạc. Tuỳ theo đặc điểm của gen chỉ thị (marker) có trong vectơ tái tổ hợp, có thể lựa chọn các tế bào tái tổ hợp theo nhiều phương pháp khác nhau. Phương pháp chọn lọc (screening) tế bào tái tổ hợp có thể sử dụng nhiều kỹ thuật khác nhau: sử dụng chỉ thị kháng sinh, sử dụng chỉ thị màu hoặc lai phân tử với mẫu dò đánh dấu phóng xạ, huỳnh quang...

5. 1. Chọn lọc dòng tế bào tái tổ hợp bằng chỉ thị kháng sinh

Chọn lọc dòng tái tổ hợp bằng chỉ thị kháng sinh là phương pháp đơn giản và dễ sử dụng. Do hiệu quả biến nạp thường rất thấp, vì vậy sau khi biến nạp cần phải chọn lọc đúng các tế bào mang vectơ tái tổ hợp.

Dựa vào gen kháng chất kháng sinh trong vectơ tái tổ hợp, chuẩn bị môi trường dinh dưỡng có bổ sung chất kháng sinh tương ứng để nuôi cấy các tế bào sau biến nạp. Trong môi trường có chất kháng sinh, các tế bào chứa vectơ tái tổ hợp có gen kháng chất kháng sinh tồn tại và phát triển được, những tế bào không có vectơ tái tổ hợp không phát triển được. Ví dụ, các vectơ tách dòng mang gen kháng ampicilin -  $am^R$  (pBR322, pUC 18...) có thể bổ sung một lượng ampicilin thích hợp vào môi trường nuôi cấy vi khuẩn sau biến nạp để chọn các dòng tái tổ hợp.

Phương pháp dùng chỉ thị kháng sinh thường có hiệu quả không cao, do một số tế bào không mang vectơ tái tổ hợp, nhưng do xuất hiện các đột biến kháng chất kháng sinh, nên vẫn có thể phát triển được. Để đảm bảo hiệu quả sàng lọc chuẩn xác, thường kết hợp sử dụng chỉ thị hai loại kháng sinh, hoặc đồng thời sử dụng gen chỉ thị kháng sinh cùng với các gen chỉ thị màu, chỉ thị enzym...

### 5. 2. Chọn lọc dòng tế bào tái tổ hợp bằng gen chỉ thị màu

Sử dụng chỉ gen chỉ thị màu kết hợp với gen chỉ thị kháng sinh là phương pháp chọn lọc dòng tái tổ hợp có hiệu quả cao. Trong các kỹ thuật sàng lọc thể tái tổ hợp ở tế bào vi khuẩn, gen *LacZ* kết hợp với IPTG và X-gal làm gen chỉ thị màu được sử dụng tương đối phổ biến. Gen *LacZ* là gen chỉ thị có giá trị cao, dễ nhận biết dòng tái tổ hợp nhờ màu sắc của khuẩn lạc, được sử dụng nhiều trong kỹ thuật gen.

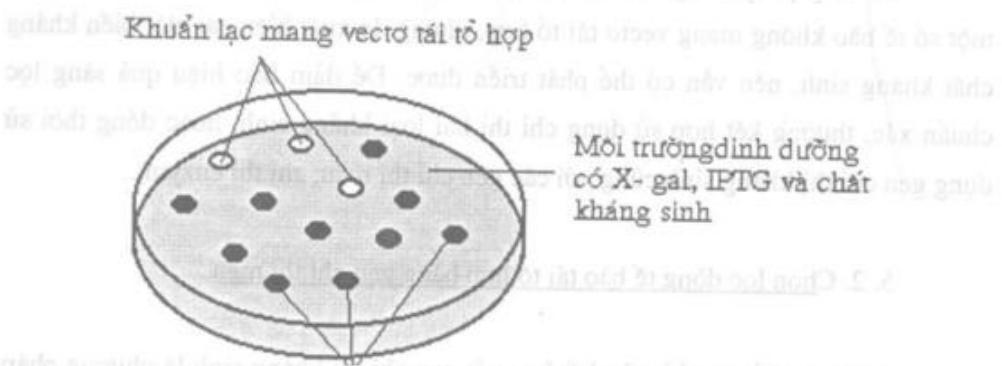
Gen *LacZ* mã hoá enzym  $\beta$ -galactosidase của vi khuẩn *E. coli*, khi có chất cảm ứng IPTG (isopropyl thiogalactoside - chất đồng đẳng của đường lactose) enzym  $\beta$ -galactosidase được tổng hợp trong tế bào *E. coli*.

Nếu môi trường có mặt X-gal (5-bromo-4-cloro-3-indolyl- $\beta$ -D-galactopyranoside), enzym  $\beta$ -galactosidase có khả năng thuỷ phân X-gal từ một hợp chất không màu thành một chất có màu xanh.

Do đó, ở môi trường nuôi cấy vi khuẩn sau khi biến nạp có bổ sung X-gal và chất cảm ứng IPTG, các tế bào vi khuẩn có gen *LacZ* nguyên vẹn (không tạo

vector tái tổ hợp) tổng hợp enzym  $\beta$ - galactosidase phân huỷ X-gal, các tế bào này phát triển thành khuẩn lạc có màu xanh.

Nếu đoạn cài DNA được gắn vào giữa gen LacZ (tạo vector tái tổ hợp) làm cho gen LacZ mất hoạt tính, enzym  $\beta$ -galactosidase không được tổng hợp, tế bào vi khuẩn phát triển thành các khuẩn lạc có màu trắng. Lựa chọn các khuẩn lạc có màu trắng, cấy truyền vào môi trường dinh dưỡng có thạch, thu được dòng tế bào mang vector tái tổ hợp.



Khuẩn lạc không có vector tái tổ hợp

**Hình 4.16. Sàng lọc dòng tái tổ hợp bằng màu sắc khuẩn lạc**

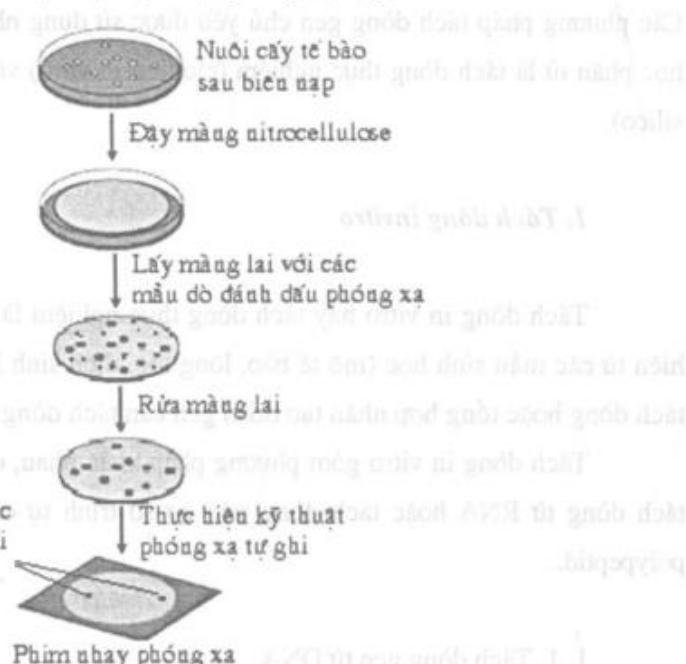
### 5.3. Chon lọc dòng tế bào tái tổ hợp bằng phương pháp lai phân tử

Phương pháp lai phân tử là một trong những phương pháp chọn lọc dòng tái tổ hợp có độ chính xác cao. Tuy nhiên, kỹ thuật thực hiện tương đối phức tạp, tốn kém về hoá chất và đòi hỏi các thiết bị chuyên dùng.

Tùy theo điều kiện nghiên cứu, có thể sử dụng phương pháp lai acid nucleic (DNA, RNA) với mẫu dò đánh dấu phong xạ hoặc đánh dấu huỳnh quang (mẫu dò là các đoạn oligo nucleotid được đánh dấu phong xạ, hoặc huỳnh quang). Các bước thực hiện chủ yếu gồm:

- Sau khi biến nạp, các tế bào vi khuẩn (hoặc nấm men) được nuôi cấy ở hộp Petri trên môi trường có 2% agar, để mỗi tế bào phát triển thành một khuẩn lạc. Đặt một màng lai (nitrocellulose) lên trên hộp Petri, để các khuẩn lạc in dấu lên màng lai ở các vị trí tương ứng.
- Lấy màng lai đã in dấu các khuẩn lạc vi khuẩn sau biến nạp, đem lai phân tử với mẫu dò đánh dấu phóng xạ (hoặc huỳnh quang), mẫu dò phải có trình tự tương ứng với đoạn đặc hiệu của gen tái tổ hợp.

- Ủ ở nhiệt độ thích hợp để tạo phản ứng lai. Rửa màng lai để loại bỏ các mẫu dò không được lai và thực hiện kỹ thuật phóng xạ tự ghi.



Hình 4.17. Sàng lọc dòng tái tổ hợp bằng kỹ thuật lai phân tử

Kết quả hiện hình phóng xạ cho thấy ở những vị trí có phản ứng lai sẽ có các chấm đen trên phim nhạy phóng xạ, hoặc các vết màu khi hiện hình huỳnh quang. Lựa chọn các khuẩn lạc ở các vị trí tương ứng trong hộp Petri thu được các dòng tái tổ hợp.

### **III. Các phương pháp tách dòng gen**

Tách dòng gen (gene cloning), còn được gọi với nhiều tên khác nhau: tạo dòng gen, nhân dòng gen, phân lập gen... Tách dòng gen là tập hợp các kỹ thuật nhằm đưa một gen, một đoạn DNA cần thiết (DNA insert) vào tế bào chủ (host cell), tạo điều kiện thích hợp để các tế bào chủ phân chia, tạo nên vô số các tế bào cùng mang một đoạn DNA insert giống nhau, tạo nên một dòng tế bào tái tổ hợp mang gen cần tách dòng.

Tách dòng gen có thể được thực hiện theo nhiều phương pháp khác nhau. Các phương pháp tách dòng gen chủ yếu được sử dụng nhiều trong công nghệ sinh học phân tử là tách dòng thực nghiệm (cloning in vitro) và tách dòng ảo (cloning in silico).

#### ***1. Tách dòng invitro***

Tách dòng in vitro hay tách dòng thực nghiệm là tách dòng gen được thực hiện từ các mẫu sinh học (mô tế bào, lông tóc, dịch sinh học...), mang các gen cần tách dòng hoặc tổng hợp nhân tạo đoạn gen cần tách dòng.

Tách dòng in vitro gồm phương pháp khác nhau, có thể tách dòng từ DNA, tách dòng từ RNA hoặc tách dòng trên cơ sở trình tự các acid amin trong chuỗi polypeptid...

##### **1. 1. Tách dòng gen từ DNA**

Tách dòng từ DNA gồm các bước cơ bản sau:

Bước 1: Chuẩn bị gen cần tách dòng (gen cần thiết, gen đích - target gene) và chọn vectơ tách dòng

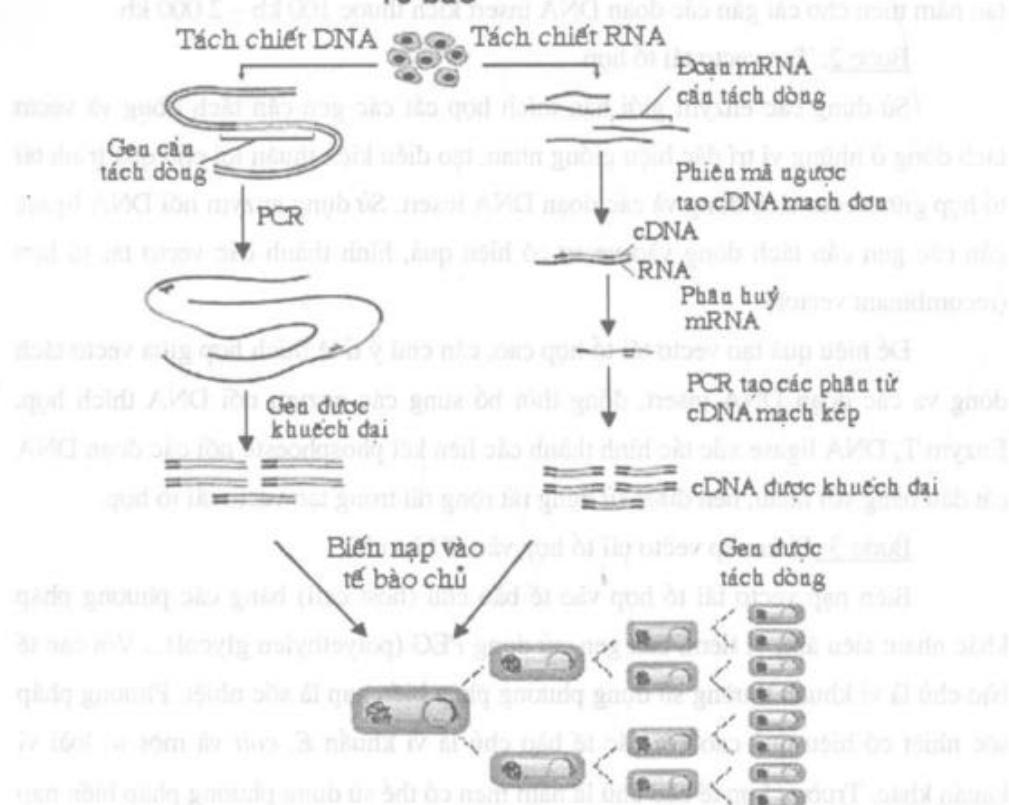
- *Chuẩn bị gen cần tách dòng:*

Trong các phòng thí nghiệm sinh học phân tử gen cần tách dòng thường được tách chiết từ các mẫu sinh học. Phương pháp chuẩn bị gen tách dòng bằng tổng hợp

nhân tạo từ các oligonucleotid ít được sử dụng, do kỹ thuật phức tạp, tốn kém và độ chuẩn xác không cao.

Trường hợp chưa biết trình tự của gen cần tách dòng, phải dựa vào kích thước của gen cần tách dòng để chọn lọc đoạn gen cần tách dòng. Các bước thực hiện chủ yếu: tách chiết DNA bộ gen, nhân gen bằng PCR, sử dụng enzym giới hạn cắt thành các đoạn DNA nhỏ, điện di trên gel agar (hoặc gel polyacrylamid) với thang DNA chuẩn. Từ kết quả điện di, dựa vào kích thước của gen cần tách dòng để lựa chọn được đoạn DNA chứa gen cần tách dòng.

### Tế bào



Hình 4.18. Sơ đồ kỹ thuật tách dòng thực nghiệm

Khi biết trình tự của gen cần tách dòng, có thể dựa vào các trình tự bảo thủ của gen cần tách dòng để thiết kế cặp mồi (primer) đặc hiệu cho phản ứng PCR.

Sau khi tách chiết và tinh sạch DNA nguồn, thực hiện kỹ thuật nhân gen PCR với cặp mồi đã thiết kế để “câu” gen cần tách dòng.

- *Chọn vectơ tách dòng:*

Dựa vào kích thước gen cần tách dòng và mục đích tách dòng có thể chọn vectơ tách dòng là plasmid, phage, các nhiễm sắc thể nhân tạo...

Vectơ tách dòng là plasmid phù hợp với tế bào chủ là tế bào vi khuẩn, có thể cho gắn đoạn DNA insert kích thước khoảng 3 - 10 kb. Vectơ tách dòng là phage cho cài gắn đoạn DNA từ 15 - 30 kb vẫn hoạt động bình thường, thích hợp với các tế bào chủ là vi khuẩn, tế bào động vật, thực vật. Vectơ tách dòng là các nhiễm sắc thể nhân tạo nấm men cho cài gắn các đoạn DNA insert kích thước 100 kb – 2 000 kb.

**Bước 2: Tạo vectơ tái tổ hợp**

Sử dụng các enzym giới hạn thích hợp cắt các gen cần tách dòng và vectơ tách dòng ở những vị trí đặc hiệu giống nhau, tạo điều kiện thuận lợi cho quá trình tái tổ hợp giữa vectơ tách dòng và các đoạn DNA insert. Sử dụng enzym nối DNA ligase gắn các gen cần tách dòng vào vectơ có hiệu quả, hình thành các vectơ tái tổ hợp (recombinant vector).

Để hiệu quả tạo vectơ tái tổ hợp cao, cần chú ý tỉ lệ thích hợp giữa vectơ tách dòng và các đoạn DNA insert, đồng thời bổ sung các enzym nối DNA thích hợp. Enzym T<sub>4</sub> DNA ligase xúc tác hình thành các liên kết phosphoeste nối các đoạn DNA cắt đầu bằng với nhau, nên được sử dụng rất rộng rãi trong tạo vectơ tái tổ hợp.

**Bước 3: Biến nạp vectơ tái tổ hợp vào tế bào chủ**

Biến nạp vectơ tái tổ hợp vào tế bào chủ (host cell) bằng các phương pháp khác nhau: siêu âm, vi tiêm, bắn gen, sử dụng PEG (polyethylen glycol)... Với các tế bào chủ là vi khuẩn thường sử dụng phương pháp biến nạp là sốc nhiệt. Phương pháp sốc nhiệt có hiệu quả cao với các tế bào chủ là vi khuẩn *E. coli* và một số loài vi khuẩn khác. Trường hợp tế bào chủ là nấm men có thể sử dụng phương pháp biến nạp là xung điện, bắn gen...

Tế bào mang vectơ tái tổ hợp còn được gọi là tế bào tái tổ hợp hay thể tái tổ hợp. Số lượng tế bào tái tổ hợp tuỳ thuộc hiệu quả biến nạp. Thông thường sốc nhiệt có hiệu quả 1/ 10 000 đến 1/ 100 000, xung điện đạt hiệu quả tương đối cao (1/ 100).

Trong các tế bào chủ vectơ tái tổ hợp được nhân lên, tạo một số lượng bản sao nhất định.

**Bước 4:** Sàng lọc các dòng tái tổ hợp

Sàng lọc (screening) các thể tái tổ hợp nhằm chọn lọc các tế bào mang vectơ tái tổ hợp trong quần thể. Tuỳ thuộc gen chỉ thị (marker) trong các vectơ tái tổ hợp là chỉ thị kháng sinh, hoặc chỉ thị màu để lựa chọn phương pháp thích hợp. Cấy truyền các khuẩn lạc có vectơ tái tổ hợp được chọn lựa sang các ống thạch nghiêng để giữa các chủng giống, đồng thời thực hiện kỹ thuật điện di với thang DNA chuẩn hoặc giải trình tự gen, để kiểm tra chính xác dòng tái tổ hợp.

**Bước 5:** Nuôi cấy dòng tái tổ hợp thu sinh khối và protein tái tổ hợp.

Nuôi cấy các dòng tái tổ hợp trong môi trường dinh dưỡng ở các điều kiện thích hợp, khi mật độ tế bào đạt  $10^8$  -  $10^9$  tế bào / 1ml có thể li tâm thu sinh khối tế bào. Tế bào được phá bằng các phương pháp siêu âm hoặc nghiên sinh khối tế bào với bì thuỷ tinh hoặc xử lí lysozym... và lựa chọn các kỹ thuật thích hợp tách và tinh chế protein tái tổ hợp.

### 1. 2. Tách dòng gen từ RNA thông tin (mRNA)

Tách dòng gen từ mRNA là một quá trình phức tạp, có thể thực hiện theo các bước sau:

**Bước 1:** Tách chiết mRNA

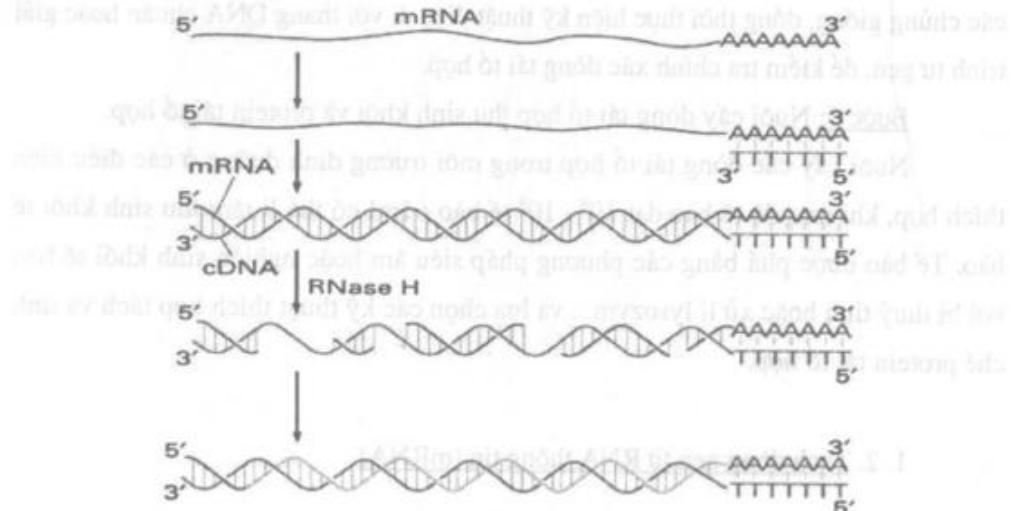
RNA thông tin được tách chiết theo các protocol thông dụng trong các phòng thí nghiệm. Có thể sử dụng sắc ký ái lực với bì từ có mang oligo nucleotid T. Các phân tử mRNA có đuôi poly A tạo liên kết với oligo - dT, làm cho các phân tử mRNA bám trên bề mặt các viên bi từ. Sử dụng kỹ thuật li tâm phân đoạn tách mRNA ra khỏi bì từ để thu mRNA. Có thể giữ mRNA ở nhiệt độ - 85°C để tiếp tục nghiên cứu.

**Bước 2:** Tổng hợp cDNA mạch kép

Sử dụng kỹ thuật PCR ngược có sự tham gia của enzym phiên mã ngược (reverse transcriptase enzyme), enzym DNA polymerase để tổng hợp cDNA (cDNA - 113.190.1.105 downloaded 60930.pdf at Sun Jul 01 23:28:18 ICT 2012

complementary DNA). Thuỷ phân sợi khuôn mRNA bằng enzym RNase H, sau đó bổ sung enzym DNA polymerase và các dNTP để tổng hợp sợi đơn cDNA thứ hai, tạo phân tử cDNA mạch kép.

Bước 3, 4, 5: Gắn cDNA mạch kép với các vectơ tách dòng thích hợp, hình thành các vectơ tái tổ hợp. Biến nạp vào tế bào chủ, sau đó sàng lọc các tế bào mang vectơ tái tổ hợp và cấy truyền vào các ống thạch nghiêng tạo nên các dòng khác nhau được tách dòng từ mRNA.



Hình 4.19. Tách dòng gen từ mRNA

Kết quả từ mRNA có thể tạo nên các dòng vi khuẩn mang đoạn cDNA mạch kép, tương ứng với các gen cần tách dòng.

## 2. Tách dòng in silico

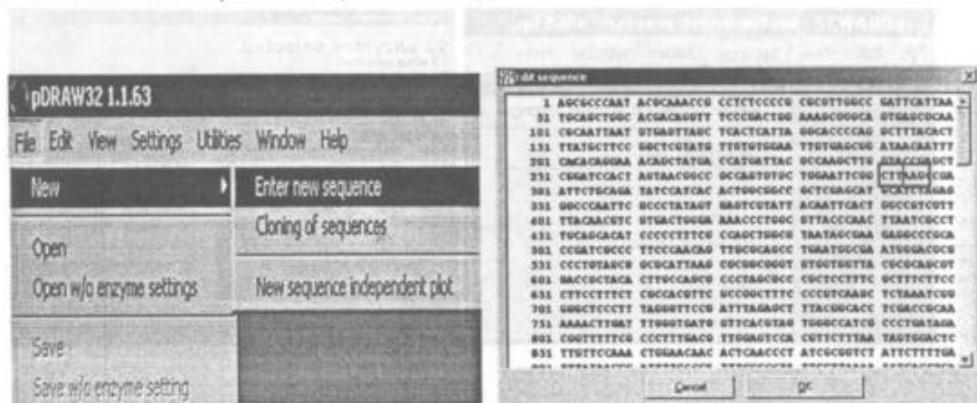
Tách dòng in silico hay tách dòng ảo (cloning in silico) là sự tổng hợp, phân tích các kết quả tách dòng gen in vitro, trên cơ sở thông tin từ các ngân hàng dữ liệu (STS, EST...) để lựa chọn một đoạn DNA hoặc một gen cần thiết nào đó. Ứng dụng thành tựu công nghệ thông tin, tách dòng in silico giúp con người tìm kiếm nhanh các

gen cần thiết, các vectơ tách dòng, xác định các điểm cắt của các enzym giới hạn hoặc thử nghiệm các mô hình hoạt động gen, hiệu quả chuyển gen...

Nguyên tắc tách dòng ảo: Vào mạng internet để lựa chọn các phần mềm phù hợp (BLAST, pDRAW32, FASTA3,...). Thực hiện các yêu cầu nghiên cứu theo các mục đích để ra như: chọn vectơ tách dòng, chọn đoạn DNA insert, chọn enzym giới hạn thích hợp, biến nạp vectơ tái tổ hợp vào tế bào chủ, hoặc thử nghiệm khả năng biểu hiện gen...

Kết quả tách dòng in silico cho phép lựa chọn phương án thiết kế vectơ tái tổ hợp có hiệu quả, dự đoán trước kết quả tách dòng và hiệu quả biểu hiện gen cũng như mức độ thành công của thực nghiệm. Ví dụ, cần thiết kế vectơ tái tổ hợp giữa đoạn DNA kích thước khoảng 958 bp, có thể thực hiện gồm các bước chủ yếu:

Bước 1: Sử dụng phần mềm pDRAW32 để chọn vectơ tách dòng. Giả thiết chọn vectơ pCR 2.1 trong các loại vectơ hiển thị, đồng thời kiểm tra trình tự nucleotid của vectơ pCR 2.1 và tìm vị trí cắt của enzym giới hạn thích hợp. Ví dụ, đoạn DNA insert có kích thước 958 bp, có thể cài gắn vào vectơ pCR 2.1 ở vị trí 294 và kết thúc ở vị trí 1252 (hình 4.20.1).



Hình 4. 20. 1. Vị trí gắn đoạn DNA insert vào vectơ pCR2.1

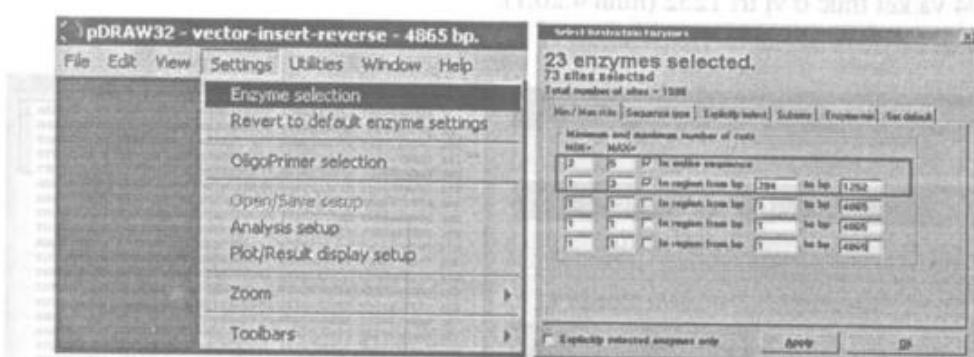
nhà khoa học

Tùy theo mục đích nghiên cứu có thể tiếp tục các lựa chọn, bằng cách kích chuột vào ô thích hợp.

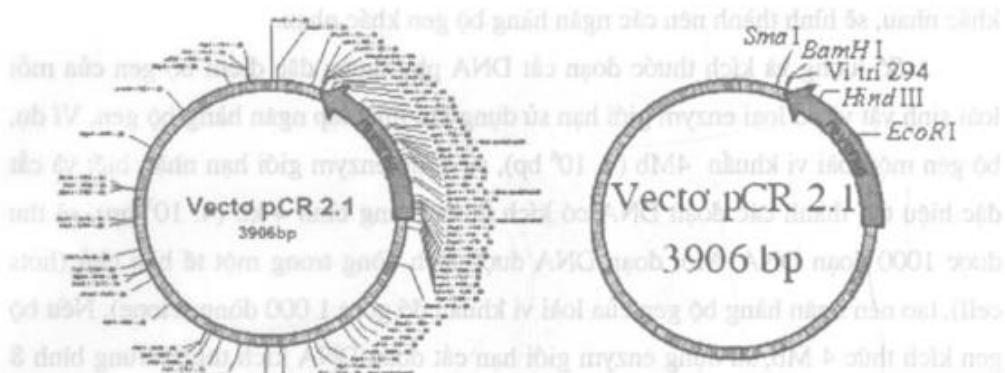
**Hình 4. 20.2. Vị trí đoạn DNA cài gắn vào vectơ pCR2.1**

Ví dụ, đoạn DNA cài (DNA insert) có thể được gắn vào vectơ tách dòng pCR 2.1 theo chiều mũi tên (hình 4. 20. 2).

Bước 2: Hiển thị các loại enzym giới hạn có thể cắt đoạn DNA cài và vectơ pCR 2.1, để có thể lựa chọn enzym phù hợp nhất.

**Hình 4. 20.3. Các loại enzym giới hạn cắt đoạn cài DNA và vectơ pCR2.1**

Hiển thị tất cả các enzym giới hạn có thể cắt vectơ tách dòng, lựa chọn enzym giới hạn phù hợp (hình 4.20.4)

**Hình 4. 20.3. Các loại enzym giới hạn được lựa chọn**

Bằng phương pháp tách dòng in silico với đoạn DNA cần tách dòng có kích thước 958 bp, có thể sử dụng vectơ tac dòng pCR2.1, với các loại enzym giới hạn *Sma I*, *BamH I*, *Hind III* và *EcoR I* có thể tạo được vectơ tái tổ hợp hiệu quả cao.

Tách dòng in silico ngày càng được ứng dụng rộng rãi trong các nghiên cứu nhằm tạo nên các sản phẩm theo ý muốn con người, sử dụng trong nghiên cứu đột biến gen, hiệu quả chuyển gen hoặc chẩn đoán các rối loạn di truyền...

#### **IV. Ngân hàng bộ gen (genome bank)**

##### **1. Khái niệm ngân hàng bộ gen**

Ngân hàng bộ gen là tập hợp các đoạn DNA bộ gen được tách dòng trong tế bào chủ, tạo nên các dòng tế bào khác nhau, mỗi dòng mang một đoạn DNA khác nhau của bộ gen. Bộ gen của sinh vật bậc cao gồm các gen phân đoạn, có các exon xen kẽ intron. Do đó, ngân hàng bộ gen bao gồm các đoạn DNA có các trình tự mang mã di truyền (exon) xen kẽ với các trình tự không mang mã di truyền (intron).

Mỗi loại enzym giới hạn cắt đặc hiệu DNA ở những trình tự nhất định. Một bộ gen, khi sử dụng các loại enzym giới hạn khác nhau, tạo nên số lượng đoạn cắt

khác nhau, sẽ hình thành nên các ngân hàng bộ gen khác nhau.

Số lượng và kích thước đoạn cắt DNA phụ thuộc đặc điểm bộ gen của mỗi loài sinh vật và số loại enzym giới hạn sử dụng khi thiết lập ngân hàng bộ gen. Ví dụ, bộ gen một loài vi khuẩn 4Mb ( $4 \cdot 10^6$  bp), sử dụng enzym giới hạn nhận biết và cắt đặc hiệu tạo thành các đoạn DNA có kích thước trung bình 4 kb ( $4 \cdot 10^3$  bp), sẽ thu được 1000 đoạn DNA. Mỗi đoạn DNA được tách dòng trong một tế bào chủ (host cell), tạo nên ngân hàng bộ gen của loài vi khuẩn đó gồm 1 000 dòng (clone). Nếu bộ gen kích thước 4 Mb, sử dụng enzym giới hạn cắt đoạn DNA kích thước trung bình 8 kb, ngân hàng bộ gen được thiết lập gồm 500 dòng...

Ngân hàng bộ gen, có vai trò quan trọng trong các nghiên cứu cấu trúc, chức năng gen và bộ gen và giải trình tự bộ gen...

## **2. Thiết lập ngân hàng bộ gen**

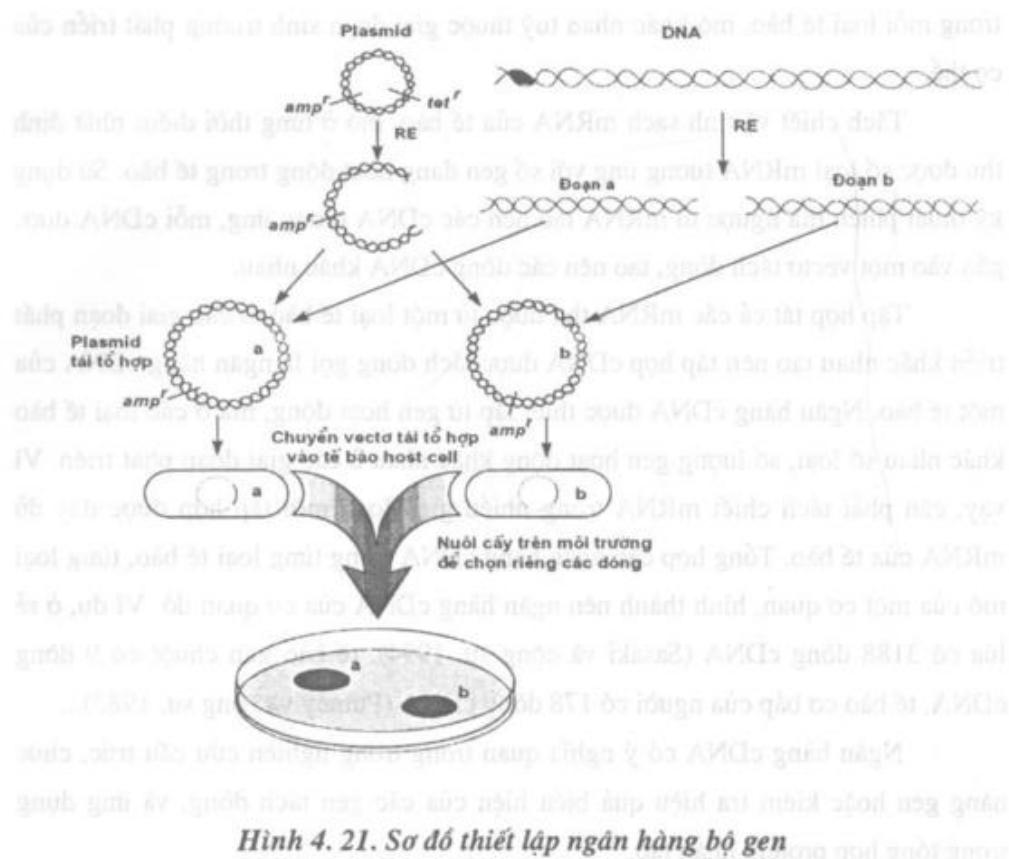
Thiết lập ngân hàng bộ gen có thể thực hiện ở các mức độ khác nhau: ngân hàng gen của một nhiễm sắc thể, một tế bào hay một loài.

Sử dụng các loại enzym giới hạn khác nhau, tạo số lượng các đoạn cắt DNA với kích thước khác nhau, hình thành các ngân hàng bộ gen có số dòng khác nhau. Thiết lập ngân hàng bộ gen phải đảm bảo ngân hàng bộ gen chứa đầy đủ hàm lượng DNA của một nhiễm sắc thể, của tế bào hoặc của loài.

Thiết lập ngân hàng bộ gen có thể thực hiện theo các bước cơ bản sau:

Tách chiết và tinh sạch DNA bộ gen, sử dụng một hoặc một số loại enzym giới hạn cắt DNA bộ gen thành từng đoạn có kích thước nhỏ theo ý muốn. Chú ý tính toán vị trí cắt và sử dụng enzym giới hạn thích hợp, để mỗi đoạn cắt có ít nhất một gen.

Tách riêng từng đoạn DNA và gắn với một vectơ tách dòng tạo nên các vectơ tái tổ hợp, mỗi vectơ tái tổ hợp được chuyển vào một tế bào chủ tạo nên một dòng.



Ngân hàng bộ gen có vai trò quan trọng trong giải trình tự gen người, động thực vật bậc cao... hoặc lưu giữ nguồn gen của các sinh vật quý hiếm, xác định vốn gen (genopool) của các quần thể, các loài sinh vật...

## V. Ngân hàng cDNA

Bộ gen sinh vật *eukaryote* có số lượng gen rất lớn, tuy nhiên số lượng gen mã hoá protein chiếm một phần nhỏ hàm lượng DNA trong tế bào (ở người khoảng 1,7 %). Trong số các gen mã hoá protein, ở mỗi loại tế bào, mỗi loại mô chỉ có một số gen nhất định hoạt động và biểu hiện. Thời gian hoạt động và biểu hiện của gen

trong mỗi loại tế bào, mô khác nhau tùy thuộc giai đoạn sinh trưởng phát triển của cơ thể.

Tách chiết và tinh sạch mRNA của tế bào, mô ở từng thời điểm nhất định thu được số loại mRNA tương ứng với số gen đang hoạt động trong tế bào. Sử dụng kỹ thuật phiên mã ngược từ mRNA tạo nên các cDNA tương ứng, mỗi cDNA được gắn vào một vectơ tách dòng, tạo nên các dòng cDNA khác nhau.

Tập hợp tất cả các mRNA thu được từ một loại tế bào, ở các giai đoạn phát triển khác nhau tạo nên tập hợp cDNA được tách dòng gọi là ngân hàng cDNA của một tế bào. Ngân hàng cDNA được thiết lập từ gen hoạt động, mà ở các loại tế bào khác nhau số loại, số lượng gen hoạt động khác nhau ở các giai đoạn phát triển. Vì vậy, cần phải tách chiết mRNA trong nhiều giai đoạn mới tập hợp được đầy đủ mRNA của tế bào. Tổng hợp các ngân hàng cDNA trong từng loại tế bào, từng loại mô của một cơ quan, hình thành nên ngân hàng cDNA của cơ quan đó. Ví dụ, ở rễ lúa có 3188 dòng cDNA (Sasaki và cộng sự, 1994), tế bào gan chuột có 9 dòng cDNA, tế bào cơ bắp của người có 178 dòng cDNA (Putney và cộng sự, 1983)...

Ngân hàng cDNA có ý nghĩa quan trọng trong nghiên cứu cấu trúc, chức năng gen hoặc kiểm tra hiệu quả biểu hiện của các gen tách dòng, và ứng dụng trong tổng hợp protein nhân tạo.

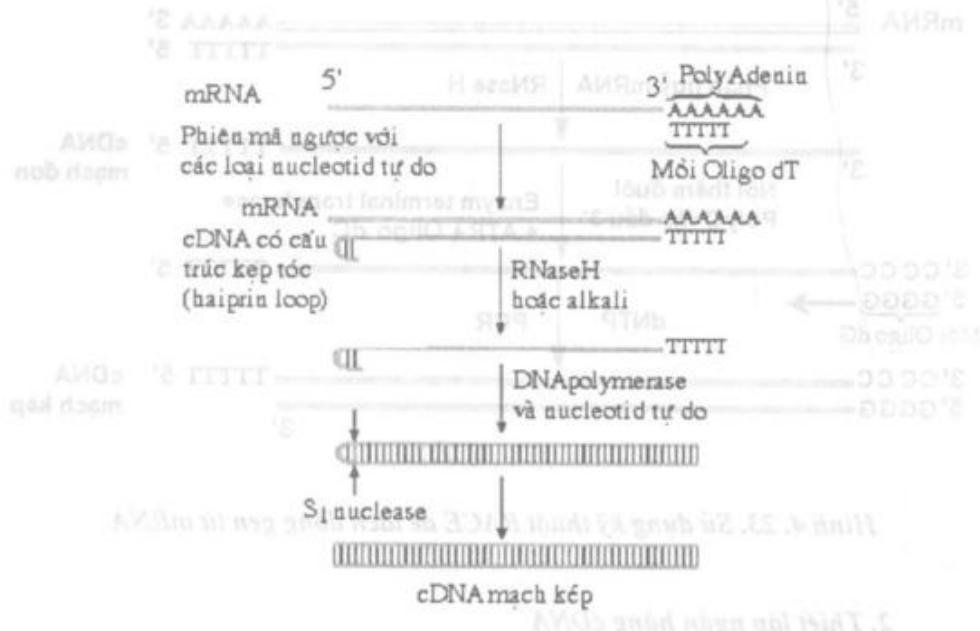
### ***1. Kỹ thuật phiên mã ngược tạo cDNA***

Kỹ thuật phiên mã ngược tạo cDNA khác nhau, gồm hai nhóm phương pháp chủ yếu:

#### **1. 1. Phiên mã ngược bằng phân huỷ hoàn toàn sợi khuôn mRNA**

Sử dụng kỹ thuật PCR với các mồi đặc hiệu có đuôi oligo - dT, các loại nucleotid tự do và enzym DNA polymerase, từ sợi khuôn mRNA tổng hợp sợi đơn cDNA có cấu trúc kẹp tóc (hairpin loop) ở đầu cuối (đầu 3'). Sử dụng enzym RNase H để phân huỷ hoàn toàn sợi khuôn mRNA, đồng thời bổ sung enzym DNA

polymerase và các nucleotid tự do để tổng hợp sợi đơn cDNA thứ 2, nhờ kéo dài cấu trúc kép tóc (cấu trúc vòng). Xử lý enzym S<sub>1</sub> nuclease để cắt bỏ đoạn cấu trúc kép tóc, tạo nên phân tử cDNA có cấu trúc xoắn kép, mang mã di truyền tương ứng với gen mã hoá mRNA.

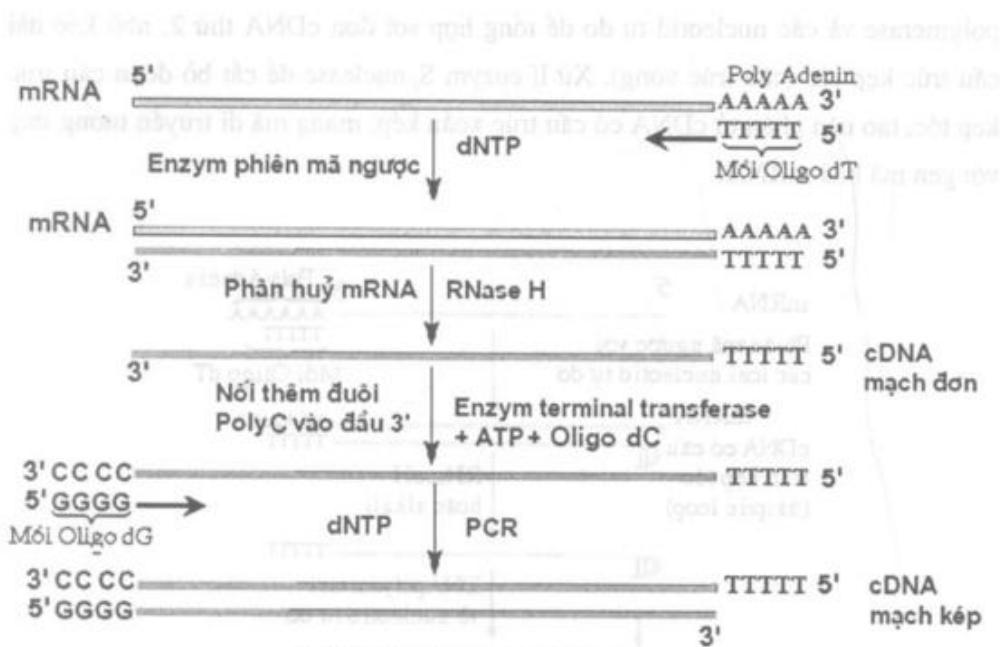


**Hình 4.22. Sơ đồ tạo cDNA theo cách phân huỷ hoàn toàn mRNA**

### 1. 2. Phiên mã ngược tạo cDNA theo kỹ thuật RACE

Kỹ thuật RACE (rapid amplification of cDNA ends - kỹ thuật nhân nhanh đầu cuối cDNA) tạo cDNA từ mRNA. Kỹ thuật RACE gồm các bước tương tự như tách dòng từ mRNA.

Sau khi tổng hợp được cDNA mạch đơn, thuỷ phân toàn bộ phân tử mRNA bằng RNase H. Thực hiện kỹ thuật nối thêm đuôi poly C (CCCC) vào đầu 3' của cDNA nhờ enzym terminal transferase và ATP. Bổ sung thêm các mồi oligo dG (GGGG), dNTP, enzym DNA polymerase và hực hiện nhân gen PCR để tổng hợp mạch cDNA thứ hai.



Hình 4. 23. Sử dụng kỹ thuật RACE để tách dòng gen từ mRNA

## 2. Thiết lập ngân hàng cDNA

Ngân hàng cDNA có thể xây dựng trên một nhiễm sắc thể, một cơ quan hoặc toàn bộ cơ thể. Do các gen hoạt động và biểu hiện ở các thời điểm khác nhau nên phải thực hiện tách chiết mRNA ở nhiều giai đoạn khác nhau. Sử dụng nhiều kỹ thuật tách chiết và tinh sạch mRNA để thu được đầy đủ mRNA, đảm bảo tính đầy đủ và chính xác của ngân hàng cDNA.

Từ mỗi phân tử mRNA thu được, thực hiện phiến mã ngược tạo nên một cDNA mạch kép. Mỗi phân tử cDNA mạch kép được gắn vào một vectơ tách dòng và chuyển vào một tế bào chủ tạo nên các dòng cDNA.

**Tài liệu tham khảo chính**

1. Dale J. W., Schantz M. (2002). From Genes to Genomes: Concepts and Applications of DNA Technology. John Wiley & Sons, Ltd.
2. Goldman N. (2001). Molecular phylogenetics: State-of-the-art methods for looking into the past, Trends in Genetics 17, 262–272
3. Griffiths A. J. F., Gelbart W.M., Lewontin R. C., Miller J.H. (2002). Modern genetic analysis. 2nd edition. W.H. Freeman, New York.
4. Hartl D. L., Jones E. W. (1999). Essential Genetics. Bartlett Pub., USA
5. Hartwell L. H., Hood L. et al. (2004). Genetics: From genes to genomes, 2nd edition. McGraw Hill, New York.
6. Hồ Huỳnh Thuỳ Dương (1988). Sinh học phân tử. NXB Giáo dục
7. Khuất Hữu Thanh (2003). Cơ sở di truyền phân tử và kỹ thuật gen. NXB Khoa học và Kỹ thuật, Hà Nội.
8. Weaver R. F., (2002). Molecular Biology - International Edition
9. Watson J. D. et. al. (2004) Molecular Biology of the Gene. FIFTH EDITION., Pearson Education, Inc., Benjamin Cummings.
10. <http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/>
11. <http://mbgd.genome.ad.jp>
12. <http://ecoli.aist-nara.ac.jp>
13. <http://bmb.med.miami.edu/ecogene>
14. <http://www.proteome.com/databases>

## CHƯƠNG 5

### **PHƯƠNG PHÁP PCR VÀ CÁC KỸ THUẬT CƠ BẢN TRONG CÔNG NGHỆ SINH HỌC PHÂN TỬ**

#### **I. Phương pháp PCR (Polymerase Chain Reaction)**

Phương pháp PCR do Kary Mullis và cộng sự phát minh, công bố tháng 10 năm 1985, tạo bước nhảy vọt của sinh học phân tử và kỹ thuật gen. Phương pháp PCR cho phép khuếch đại, tạo một số lượng rất lớn bản sao của gen (hay một đoạn DNA) trong một thời gian ngắn. Nguyên tắc PCR dựa trên cơ sở tính chất biến tính, hồi tính của DNA và nguyên lý tổng hợp DNA. Trên cơ sở trình tự của đoạn DNA khuôn (DNA khuôn, DNA đích - target sequence), đoạn mồi (primer), các nucleotid tự do (dNTP) và enzym DNA polymerase có thể tổng hợp được đoạn DNA giới hạn bởi các đoạn mồi. Lặp lại nhiều lần các chu kì nhân gen, trong một thời gian ngắn số lượng bản sao DNA tạo thành tăng theo cấp số nhân.

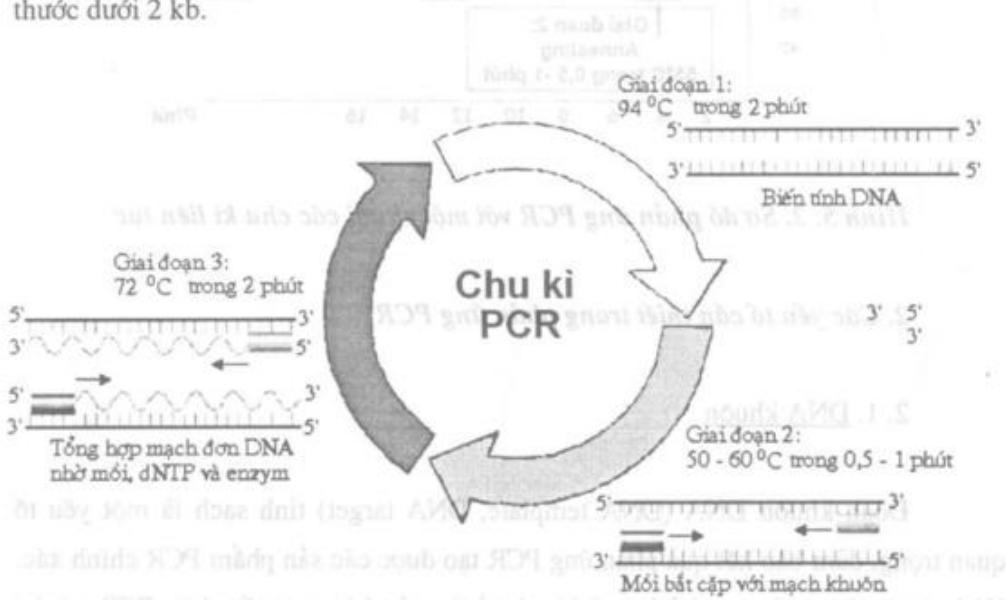
##### **1. Nguyên lý chung của phương pháp PCR**

PCR là một chuỗi phản ứng liên tục, gồm nhiều chu kỳ (cycle) kế tiếp nhau, mỗi chu kỳ gồm ba giai đoạn :

Giai đoạn biến tính (Denaturation): ở nhiệt độ cao  $90 - 95^{\circ}\text{C}$  (cao hơn  $T_m$  của DNA khuôn), làm đứt các liên kết hydro của phân tử DNA, hai mạch phân tử DNA tách rời nhau. Đoạn khuôn DNA có các đoạn dài gồm nhiều nucleotid giống nhau, hoặc tỉ lệ G - C càng cao, có nhiệt độ biến tính ( $T_m$ ) cao hơn. Do vậy cần cẩn cứ đặc điểm của DNA khuôn để lựa nhiệt độ biến tính phù hợp. Giai đoạn này kéo dài 1 phút đến 2 phút.

Giai đoạn gắn mồi (Annealing): ở nhiệt độ khoảng 55 - 65°C để các mồi bắt cặp với các mạch đơn DNA khuôn ở các đầu 3' theo nguyên lý Chargaff. Giai đoạn này khoảng 30 - 60 giây.

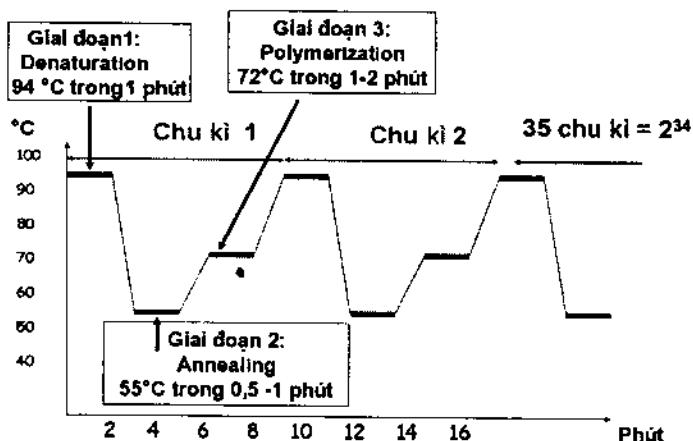
Giai đoạn tổng hợp (Elongation): nhiệt độ 70 - 72°C thích hợp với điều kiện hoạt động của enzym DNA polymerase. Enzym DNA polymerase xúc tác hoạt động tổng hợp gắn thêm các nucleotid vào cuối đoạn mồi, các mồi được kéo dài trên cơ sở bắt cặp với mạch khuôn theo nguyên lý Chargaff, tạo nên các mạch đơn DNA mới, giai đoạn này còn gọi là giai đoạn polymer hoá (polymerization). Thời gian giai đoạn này từ 30 giây đến vài chục phút, tùy thuộc vào kích thước của đoạn DNA. Thông thường, với thời gian 2 phút, tổng hợp được những đoạn DNA kích thước dưới 2 kb.



Hình 5.1. Các giai đoạn của một chu kỳ PCR

Một phản ứng PCR gồm nhiều chu kỳ liên tục, sản phẩm tạo ra ở chu kỳ trước lại được làm khuôn ở chu kỳ kế tiếp, nên số lượng bản sao tạo thành tăng theo cấp số nhân. Một phản ứng PCR thường thực hiện 20 - 40 chu kỳ, từ mỗi đoạn DNA khuôn có thể tạo nên  $2^{20} - 2^{40}$  bản sao DNA. Trong quá trình thực hiện phản

Ứng PCR, cần lưu ý những chu kỳ sau lượng khuôn tăng, lượng mồi và dNTP tự do giảm, enzym DNA polymerase hoạt động yếu dần... do đó, cần tính toán hàm lượng mồi dNTP, enzym để đảm bảo phản ứng PCR có kết quả tốt nhất.



Hình 5. 2. Sơ đồ phản ứng PCR với một chuỗi các chu kỳ liên tục

## 2. Các yếu tố cần thiết trong phản ứng PCR

### 2. 1. DNA khuôn

Đoạn khuôn DNA (DNA template, DNA target) tinh sạch là một yếu tố quan trọng, đảm bảo kết quả phản ứng PCR tạo được các sản phẩm PCR chính xác. Kích thước đoạn khuôn nhỏ hơn 3 kb cho kết quả nhân gen tốt nhất. PCR có thể khuếch đại được DNA từ những mẫu sinh học đang bị phân huỷ như vết máu, vết tinh dịch đã lâu ngày, tóc và xương của người chết, ...

### 2. 2. Các nucleotid tự do (dNTP)

Nucleotid tự do cần thiết cho phản ứng PCR gồm dATP, dTTP, dGTP, và dCTP. Trong mỗi phản ứng PCR cần chú ý đến tỉ lệ G-C trong đoạn khuôn để xác

định hàm lượng mỗi loại dNTP phù hợp. Nồng độ dNTP mỗi loại, thường sử dụng trong các phản ứng PCR khoảng 50 – 200 µM. Khi hàm lượng các loại dNTP tự do quá ít, tạo sản phẩm PCR ít không đủ để phát hiện, ngược lại nồng độ dNTP tự do quá cao thì phản ứng PCR khó thực hiện. Do đó, tuỳ theo kích thước đoạn gen và số chu kì PCR thực hiện để tính toán nồng độ dNTP thích hợp, tránh trường hợp để dư thừa dNTP khi kết thúc phản ứng hoặc thiếu một loại dNTP nào đó, gây ảnh hưởng đến kết quả PCR.

### 2.3. Mồi

Mồi (primer) là các đoạn oligonucleotid ngắn khoảng 14 - 35 nucleotid, có vai trò quan trọng quyết định thành công của phản ứng PCR. Để thực hiện nhân một đoạn DNA bằng phản ứng PCR, cần có một cặp mồi thích hợp. Một cặp mồi trong PCR gồm mồi xuôi F (Forward) và mồi ngược R (Revert). Mồi có vai trò tạo các nhóm 3' OH tự do, cần thiết cho các phản ứng polyme hoá. Mồi xuôi phải có trình tự tương đồng với trình tự của mạch đơn DNA không mang mã (mạch antisense), mồi xuôi bắt cặp ở đầu 3' của mạch antisense. Mồi ngược có trình tự tương đồng với trình tự của mạch DNA mang mã di truyền (mạch sense), mồi ngược bắt cặp với mạch sense ở đầu 3' của mạch sense.

Điều kiện chủ yếu của bảo đảm hiệu quả mồi là:

- Mồi không quá ngắn (nhỏ hơn 8 nucleotid) hoặc quá dài (lớn hơn 50 nucleotid), mồi quá ngắn kết quả PCR không chính xác, mồi quá dài khó đảm bảo thành công của phản ứng PCR. Khi thiết kế mồi, nên để dư 1-3 nucleotid ở đầu 5' của mồi, giúp cho đoạn DNA được nhân lên có trình tự nguyên vẹn.
- Trình tự của mồi có tính đặc hiệu cao, chỉ có thể bắt cặp ở đầu 3' của mạch khuôn. Các mồi không được bắt cặp với nhau, hoặc bắt cặp ở giữa mạch khuôn.
- Nhiệt độ nóng chảy ( $T_m$ ) của các mồi không chênh lệch nhau quá lớn.

Trong mồi phản ứng PCR, cần thiết phải tính toán nồng độ mồi đủ lớn, để

đảm bảo kết quả PCR tạo nên một lượng sản phẩm cần thiết. Nếu nồng độ mồi quá

thấp sẽ không đảm bảo đủ lượng sản phẩm PCR, ngược lại nồng độ mồi quá cao có thể gây hiệu quả PCR kém chính xác, do mồi có thể bám vào các vị trí không đặc hiệu làm giảm độ chính xác của phản ứng PCR.

#### 2. 4. Enzym DNA Polymerase

Enzym DNA polymerase là yếu tố rất quan trọng, có vai trò quyết định hiệu quả phản ứng PCR. Mỗi loại enzym DNA polymerase có đặc tính và vai trò khác nhau trong phản ứng PCR, sử dụng các loại enzym DNA polymerase khác nhau thu được các sản phẩm PCR có tính đặc hiệu khác nhau. Ví dụ, enzym Taq DNA polymerase được sử dụng phổ biến trong các phòng thí nghiệm. Taq DNA polymerase có hoạt tính cao ở 75 - 80°C, tốc độ xúc tác tổng hợp mạch đơn DNA rất cao (trung bình 130 -300 nucleotid/ 1 giây), nhưng có tỉ lệ bắt cặp sai giữa các nucleotid cao. Enzym Pfu DNA polymerase có hoạt tính ở nhiệt độ thích hợp 72 - 78°C, tốc độ xúc tác tổng hợp mạch đơn DNA chậm, khoảng 12 / 1 giây, nhưng tỉ lệ sai sót giảm đi nhiều lần so với sai sót khi sử dụng Taq DNA polymerase.

Hàm lượng enzym DNA polymerase cũng ảnh hưởng rất lớn đến hiệu quả PCR. Khi hàm lượng enzym DNA polymerase quá ít, không tạo đủ lượng sản phẩm PCR cần thiết. Ngược lại, quá nhiều enzym DNA polymerase, kết quả PCR tạo ra sản phẩm không đặc hiệu. Hàm lượng enzym phụ thuộc vào số chu kỳ của phản ứng PCR cũng như kích thước của đoạn DNA khuôn.

Do đó, trong các phản ứng PCR, tùy theo mục đích cụ thể của các nghiên cứu để lựa chọn loại enzym với hàm lượng thích hợp. Một số loại enzym DNA polymerase thường được sử dụng trong các phản ứng PCR gồm: T4 DNA polymerase, Taq DNA polymerase, Vent DNA polymerase, Tth DNA polymerase, Pfu DNA polymerase...

#### 2. 5. Dung dịch đệm cho PCR

Dung dịch đệm cho PCR (buffer for PCR) là một trong những yếu tố quan

ảnh hưởng đến chất lượng và hiệu quả của phản ứng PCR. Dung dịch đậm của phản ứng PCR cần đảm bảo thành phần các chất cần thiết cho hoạt động của enzym DNA polymerase như MgCl<sub>2</sub>, KCl, Tris... Trong dung dịch đậm, ion Mg<sup>2+</sup> là thành phần có ảnh hưởng rõ rệt đến hiệu quả phản ứng PCR, do ion Mg<sup>2+</sup> ảnh hưởng đến khả năng bắt cặp và gắn các mồi với mạch khuôn.

## 2. 6. Thiết bị thực hiện phản ứng PCR

Phản ứng PCR là một chuỗi các chu kỳ tuần hoàn nhiệt, do vậy có thể thực hiện phản ứng PCR trong các thiết bị chuyên dụng là các máy PCR.



**Hình 5.3. Máy nhân gen PCR**

Trường hợp phòng thí nghiệm không có máy PCR, có thể thực hiện phản ứng PCR trong các bể ủ nhiệt, có bộ đếm giờ với dụng cụ thông dụng (ống Eppendorf, kẹp...) trong phòng thí nghiệm.

Ngày nay, với sự phát triển của khoa học công nghệ ngày càng cao, các hãng sản xuất khác nhau ngày càng đưa ra các loại máy PCR thế hệ mới, có nhiều đặc điểm tiện lợi, cho kết quả cao, tốn ít thời gian thực hiện phản ứng...

### 3. Các ứng dụng chủ yếu của PCR

PCR ngày càng có nhiều ứng dụng rộng rãi trong công nghệ sinh học phân tử và kỹ thuật gen. Một số ứng dụng cơ bản của PCR gồm:

Sử dụng PCR trong tách dòng gen, xây dựng ngân hàng gen, lập các loại bản đồ gen và giải trình tự gen, bộ gen.

PCR được ứng dụng trong các nghiên cứu chẩn đoán sớm các bệnh nhiễm virus, vi khuẩn và các đột biến gây rối loạn di truyền

PCR được sử dụng trong tạo đột biến *in vitro*, và trong một số kỹ thuật chuyển gen nhằm tạo các giống vật nuôi cây trồng và vi sinh vật mới

PCR có vai trò quan trọng trong nghiên cứu nguồn gốc các loài sinh vật và trong phân loại phân tử.... (chương 7)

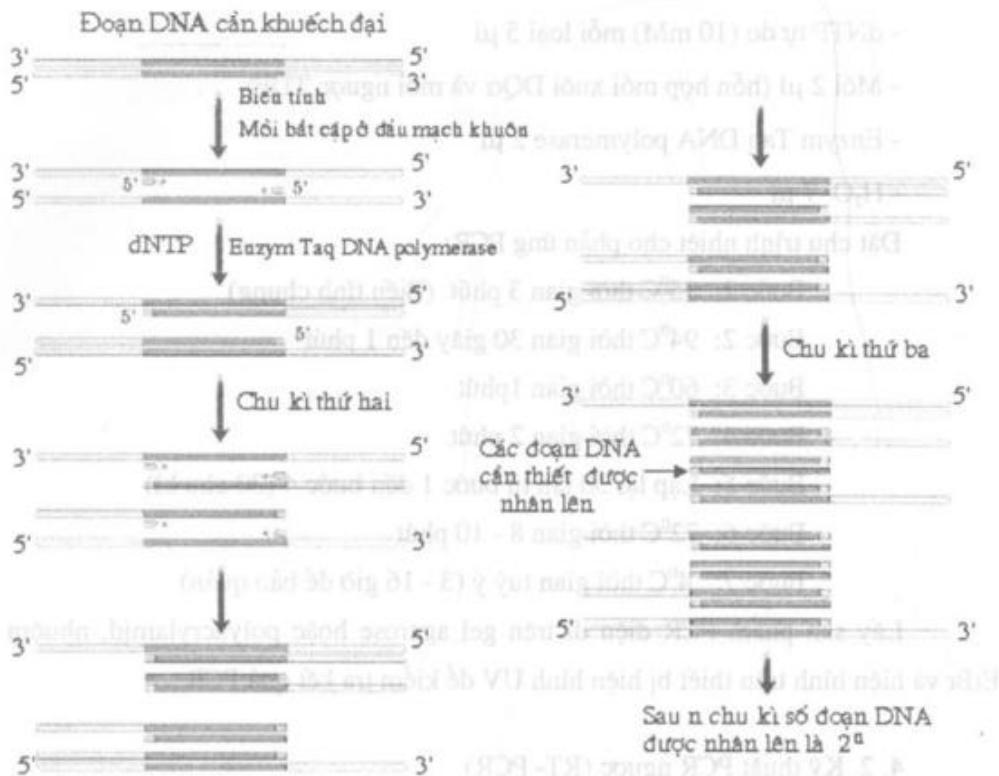
### 4. Một số kỹ thuật PCR chủ yếu

#### 4. 1. Kỹ thuật PCR chuẩn (Standard PCR)

##### a/ Nguyên lý chung

Kỹ thuật PCR chuẩn do Kary Mullis và cộng sự phát minh (1985), được sử dụng để khuếch đại một đoạn DNA. Kỹ thuật PCR chuẩn là kỹ thuật PCR cơ bản, sự cải tiến kỹ thuật PCR chuẩn tạo nên nhiều kỹ thuật PCR đặc hiệu khác nhau: PCR ngược, PCR lồng, PCR đảo, PCR phức... Các kỹ thuật PCR đặc hiệu khác nhau đều dựa trên nguyên tắc cơ bản của PCR chuẩn, một số cải tiến ở các kỹ thuật PCR đặc hiệu cho phép con người thực hiện các phản ứng nhân gen với hiệu quả cao. Kỹ thuật PCR chuẩn, với sự tham gia của các mồi đặc hiệu, các loại dNTP tự do và enzym polymerase, sau 20 -35 chu kỳ có thể nhân được số bản sao rất lớn của đoạn DNA cần khuếch đại. Kỹ thuật PCR chuẩn được thực hiện khi biết rõ các trình tự đặc hiệu ở hai đầu đoạn DNA cần nhân gen (đoạn khuôn) để có thể thiết kế các cặp mồi đặc hiệu. Trường hợp chỉ biết một trình tự đặc hiệu ở một đầu của đoạn

khuôn, chỉ có thể thiết kế được một loại mồi, phải sử dụng các kỹ thuật PCR đặc hiệu mới đảm bảo nhân gen có hiệu quả.



Hình 5.4. Sơ đồ kỹ thuật PCR chuẩn

b/ Ví dụ qui trình PCR chuẩn để kiểm tra giới tính ở người

Kỹ thuật PCR để nhân một đoạn gen trên NST Y, và một đoạn gen trên nhiễm sắc thể số 6 có thể phát hiện giới tính ở người.

Chuẩn bị 20 µl phản ứng gồm:

Cho vào ống eppendorf loại 30 µl các thành phần cơ bản của phản ứng PCR gồm:

- Dung dịch đậm 2 µl (1M TRIS-HCl pH 8,3; 500 mM KCl; 15 mM MgCl<sub>2</sub>)
- DNA khuôn: 2 µl dịch tách chiết DNA người (tách chiết từ lông tóc, máu hoặc dịch sinh học của người)
  - dNTP tự do (10 mM) mỗi loại 5 µl
  - Mỗi 2 µl (hỗn hợp mỗi xoáy DQα và mỗi ngược TDF)
  - Enzym Taq DNA polymerase 2 µl
  - H<sub>2</sub>O 7 µl

Đặt chu trình nhiệt cho phản ứng PCR:

Bước 1: 95°C thời gian 3 phút (Biến tính chung)

Bước 2: 94°C thời gian 30 giây đến 1 phút

Bước 3: 60°C thời gian 1 phút

Bước 4: 72°C thời gian 2 phút

Bước 5: Lặp lại 30 lần từ bước 1 đến bước 4 (30 chu kỳ)

Bước 6: 72°C thời gian 8 - 10 phút

Bước 7: 4°C thời gian tuỳ ý (3 - 16 giờ để bảo quản)

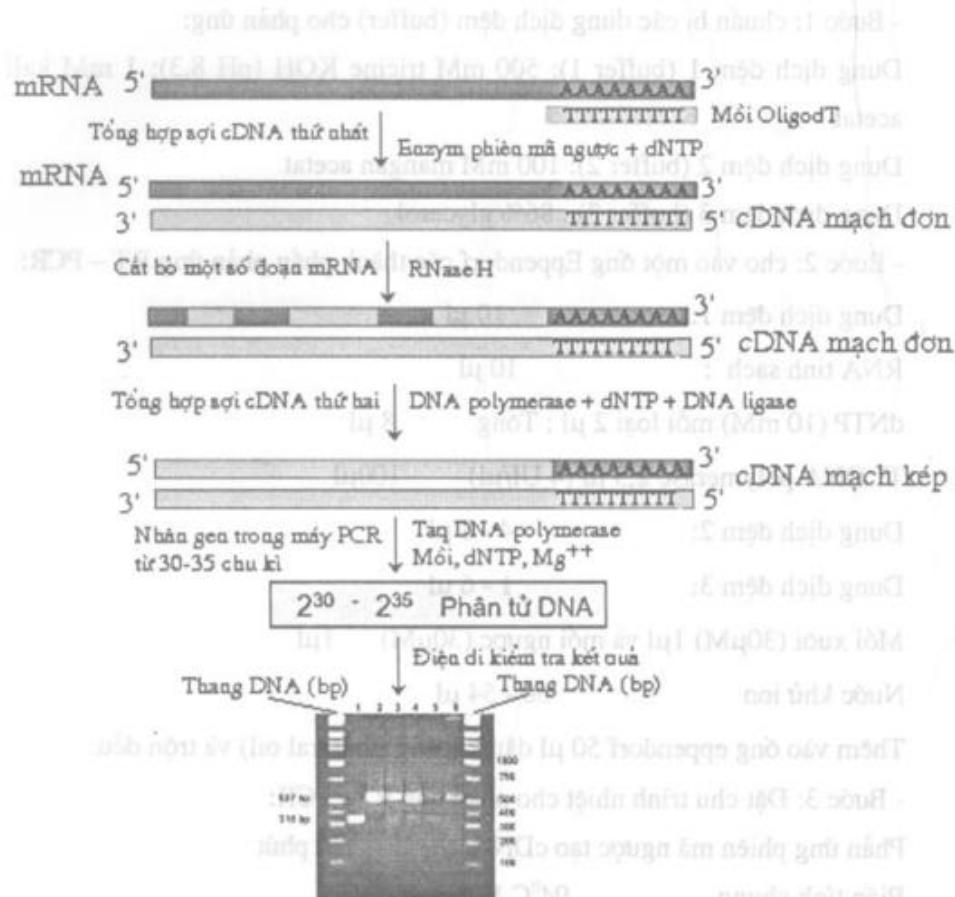
Lấy sản phẩm PCR điện di trên gel agarose hoặc polyacrylamid, nhuộm EtBr và hiện hình trên thiết bị hiện hình UV để kiểm tra kết quả PCR.

#### **4. 2. Kỹ thuật PCR ngược (RT- PCR)**

##### *a/ Nguyên lý chung*

Kỹ thuật PCR ngược (Reverse Transcriptase PCR) là kỹ thuật PCR cải tiến từ kỹ thuật PCR chuẩn, sử dụng để nhân các gen (hoặc các đoạn DNA) từ mạch khuôn là mRNA và RNA. Thực chất kỹ thuật PCR ngược gồm hai đoạn: giai đoạn phiến mã ngược từ mạch khuôn RNA tạo đoạn cDNA và giai đoạn khuếch đại đoạn DNA. Trong giai đoạn 1 sử dụng các enzym phiến mã ngược tạo nên các đoạn cDNA mạch kép, tương ứng với các đoạn khuôn là mRNA hoặc RNA. Ở giai đoạn hai, thực hiện phản ứng PCR chuẩn để nhân đoạn cDNA, tạo nên số lượng bản sao DNA cần thiết.

Kỹ thuật PCR ngược có thể thực hiện cả hai giai đoạn kế tiếp nhau trong cùng một ống eppendorf, hoặc thực hiện gián đoạn: phiên mã ngược tạo cDNA trong một ống eppendorf, sau đó tiến hành phản ứng PCR trong một ống eppendorf khác.



Hình 5.5. Sơ đồ kỹ thuật PCR ngược (RT-PCR)

b/ Ví dụ qui trình RT- PCR nhân đoạn DNA từ RNA của virus

Phản ứng PCR ngược có thể được sử dụng để kiểm tra nhanh khả năng nhiễm các loại bệnh do virus ở động vật, thực vật và con người.

Có nhiều phương pháp khác nhau được sử dụng để nhân đoạn DNA từ mạch khuôn là RNA virus, mỗi phương pháp gồm các bước cơ bản giống nhau, có thể có sự khác nhau ở thành phần hóa chất trong dung dịch đệm, hoặc sử dụng enzym phiên mã ngược có nguồn gốc khác nhau.

- Bước 1: chuẩn bị các dung dịch đệm (buffer) cho phản ứng:

Dung dịch đệm 1 (buffer 1): 500 mM tricine KOH (pH 8,3); 1 mM kali acetat

Dung dịch đệm 2 (buffer 2): 100 mM mangan acetat

Dung dịch đệm 3 (buffer 3) : 86% glycerol

- Bước 2: cho vào một ống Eppendorf các thành phần phản ứng RT – PCR:

Dung dịch đệm 1: 10 µl

RNA tinh sạch : 10 µl

dNTP (10 mM) mỗi loại 2 µl ; Tổng 8 µl

T<sup>th</sup> DNA polymerase 2,5 µl (4 UI/µl) 100µl

Dung dịch đệm 2: 4 - 8 µl

Dung dịch đệm 3: 1 - 6 µl

Mỗi xuôi (30µM) 1µl và mỗi ngược (30µM) 1µl

Nước khử ion 50 - 54 µl

Thêm vào ống eppendorf 50 µl dầu khoáng (mineral oil) và trộn đều.

- Bước 3: Đặt chu trình nhiệt cho phản ứng RT - PCR:

Phản ứng phiên mã ngược tạo cDNA 30 phút

Biến tính chung 94°C 1- 2 phút

Chu kỳ thứ nhất:

Biến tính (Denaturation) 94°C 10 giây

Gắn mồi (Annealing) 50 - 65°C 30 giây

Tổng hợp (Elongation) 72°C 90 giây

Lặp lại các bước của chu kỳ thứ nhất từ 30 -35 lần

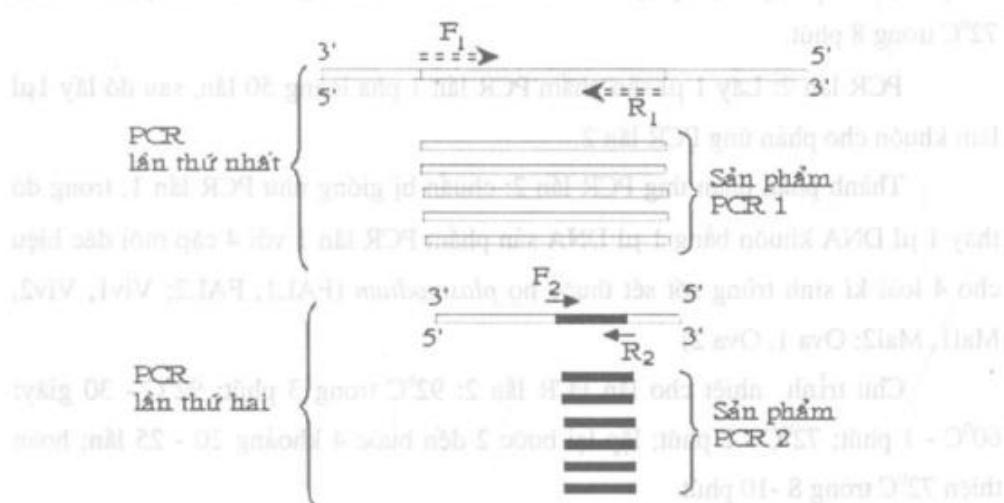
Hoàn thiện các quá trình tổng hợp 72°C 8 -10 phút

Giữ ở 4°C trong thời gian chờ xử lí kết quả (1 - 6 giờ).

#### 4. 3. Kỹ thuật PCR lồng (Nested PCR)

##### a/ Nguyên lý chung

Kỹ thuật PCR lồng (Nested PCR) là kỹ thuật thực hiện 2 lần PCR chuẩn kế tiếp nhau. Trong đó, lần PCR thứ 2 (Nested 2) sử dụng cặp mồi thứ 2 và kết quả khuếch đại các đoạn DNA của lần PCR thứ nhất (Nested 1) làm khuôn, để nhận các đoạn DNA có tính đặc hiệu cao. PCR lồng được sử dụng phổ biến trong phân loại phân tử, cho phép khuếch đại các đoạn DNA đặc trưng cho các đơn vị dưới loài, phân biệt sự khác nhau giữa các giống động thực vật, các chủng vi sinh vật...



Hình 5. 6. Sơ đồ kỹ thuật PCR lồng (Nested PCR)

b/ Ví dụ kỹ thuật PCR lồng phát hiện chính xác các loài ký sinh trùng sói rét có trong máu người

##### PCR lần 1 (Nested 1):

Tách chiết DNA từ các mẫu máu khô trên giấy thấm Whatman, theo Kain, Lannar (1995)

Thành phần phản ứng (Nested 1)

Dung dịch đậm 2  $\mu$ l

DNA khuôn 3  $\mu$ l (DNA tách chiết từ máu người)

dNTP (10 mM) mỗi loại 5  $\mu$ l

Mỗi 2  $\mu$ l (hỗn hợp mỗi xuôi và mỗi ngược đặc trưng với kí sinh trùng sốt rét *Plasmodium* (Plu5 và Plu6).

Taq DNA polymerase 1  $\mu$ l

Nước khử ion 10  $\mu$ l

Chu trình nhiệt cho lần PCR lần 1: 95°C trong 5 phút; 94°C - 1 phút; 58°C - 1 phút; 72°C - 2 phút; lặp lại bước 2 đến bước 4 khoảng 30 - 35 lần; hoàn thiện 72°C trong 8 phút.

PCR lần 2: Lấy 1  $\mu$ l sản phẩm PCR lần 1 pha loãng 50 lần, sau đó lấy 1  $\mu$ l làm khuôn cho phản ứng PCR lần 2.

Thành phần phản ứng PCR lần 2: chuẩn bị giống như PCR lần 1, trong đó thay 1  $\mu$ l DNA khuôn bằng 1  $\mu$ l DNA sản phẩm PCR lần 1 với 4 cặp mỗi đặc hiệu cho 4 loài kí sinh trùng sốt sét thuộc họ *plasmodium* (FAL1, FAL2; Viv1, Viv2, Mal1, Mal2; Ova 1, Ova 2)

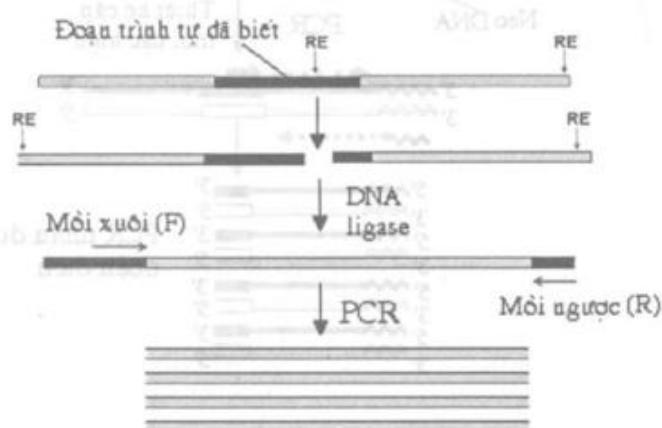
Chu trình nhiệt cho lần PCR lần 2: 92°C trong 3 phút; 92°C - 30 giây; 60°C - 1 phút; 72°C - 2 phút; lặp lại bước 2 đến bước 4 khoảng 20 - 25 lần; hoàn thiện 72°C trong 8 -10 phút.

Thu sản phẩm PCR lần 2, sau đó thực hiện điện di kết quả, nhuộm, hiện hình và xác định kết quả nhân gen.

#### 4. 4. Kỹ thuật PCR đảo (Inverse PCR)

Kỹ thuật PCR đảo là phương pháp PCR để nhân một đoạn DNA chưa biết trên phân tử DNA, nằm trước hoặc sau một đoạn DNA đã biết một trình tự. Kỹ thuật PCR đảo do Jonathan và Silver cải tiến từ PCR chuẩn năm 1989.

Các bước thực hiện kỹ thuật PCR đảo, tương tự như phương pháp PCR chuẩn. Tuy nhiên có sự khác biệt cơ bản ở giai đoạn chuẩn bị DNA khuôn (DNA template). Do chỉ biết một đoạn trình tự, nên chỉ có thể xác định được một mồi đặc hiệu, vì vậy không thể thực hiện phản ứng nhân gen bằng kỹ thuật PCR chuẩn. Để có thể thiết kế được cặp mồi đặc hiệu từ đoạn trình tự đã biết, cần sử dụng các loại enzym giới hạn (RE) cắt ở giữa đoạn trình tự đã biết (với điều kiện đoạn trình tự đã biết có một kích thước đủ lớn). Sau đó, sử dụng enzym DNA ligase nối đảo 2 đầu tạo nên đoạn DNA có các trình tự đã biết nằm ở hai đầu đoạn chưa biết. Từ các trình tự đã biết, thiết kế cặp mồi đặc hiệu, và thực hiện kỹ thuật PCR chuẩn để nhân đoạn DNA chưa biết theo ý muốn.



Hình 5. 7. Sơ đồ kỹ thuật PCR đảo (Inverse PCR)

#### 4. 5. Kỹ thuật PCR dạng neo (Anchored PCR)

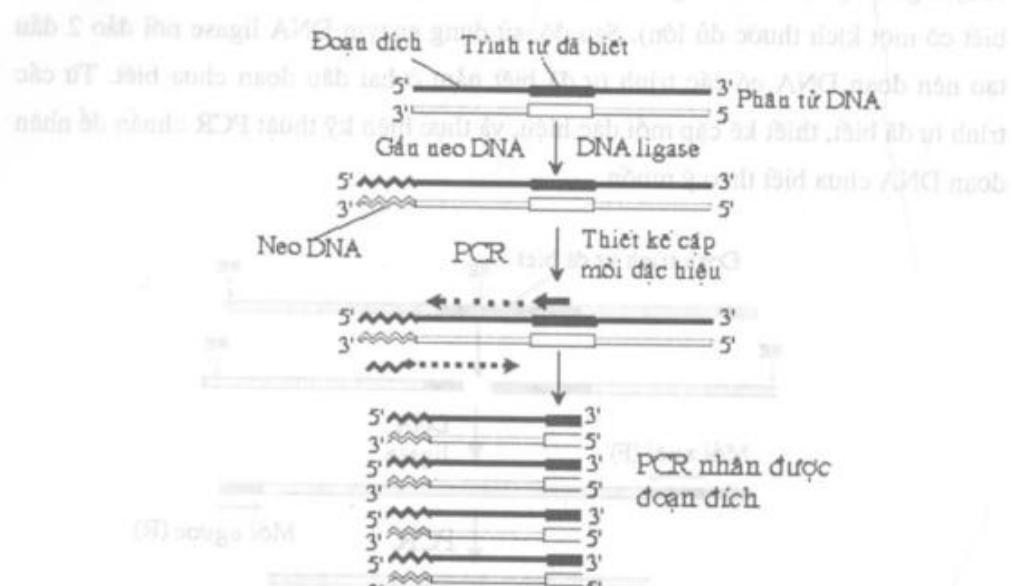
Kỹ thuật PCR dạng neo là phương pháp nhân một đoạn gen đích, khi đã biết một trình tự ngắn ở một đầu phân tử DNA, có thể thiết kế được một mồi.

Kỹ thuật PCR dạng neo là phương pháp nhân một đoạn gen đích, khi đã biết một trình tự ngắn ở một đầu phân tử DNA, có thể thiết kế được một mồi. Do trình tự đã biết có kích thước quá nhỏ, nếu sử dụng kỹ thuật PCR đảo sẽ cho hiệu quả kém.

Kỹ thuật PCR dạng neo là một trường hợp đặc biệt của PCR đảo, nhưng

không sử dụng các loại enzym giới hạn để cắt đoạn đã biết và nối đảo đầu, tạo cơ sở thiết kế cặp mồi đặc hiệu được. Người ta thường sử dụng một đoạn đặc hiệu gọi là đoạn neo (Anchored DNA) gắn vào đầu chưa biết của đoạn DNA khuôn, nhờ các adaptor đặc hiệu và enzym nối (DNA ligase).

Dựa vào trình tự đã biết ở 2 đầu để thiết kế các cặp mồi đặc hiệu, và thực hiện kỹ thuật PCR để nhân đoạn gen đích cần thiết.

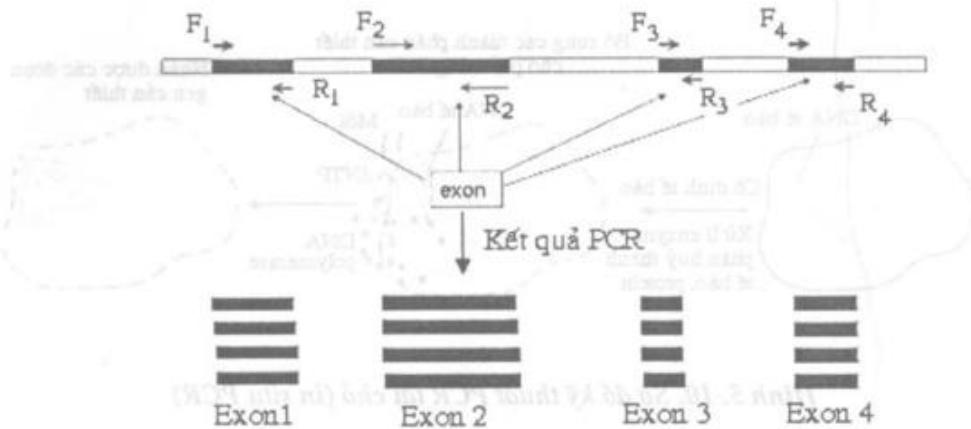


Hình 5. 8. Sơ đồ kỹ thuật PCR neo (Anchored PCR)

#### 4. 6. Kỹ thuật PCR phức (Multiplex PCR)

Kỹ thuật PCR phức là phương pháp PCR sử dụng đồng thời cùng một lúc nhiều cặp mồi đặc hiệu khác nhau, để nhân các đoạn DNA đặc trưng khác nhau trên một phân tử DNA, hoặc trên các phân tử DNA khác nhau. Kỹ thuật PCR phức thường sử dụng để nhân các exon của một gen nào đó của sinh vật bậc cao. Ví dụ, khi nghiên cứu gen chuyển hóa hypoxanthin (HPRT) có 9 exon khác nhau, có thể dùng PCR phức với 16 - 18 cặp mồi khác nhau để nhân các exon cần thiết.

Kỹ thuật PCR phức thường sử dụng phát hiện sớm tôm, cá bị nhiễm bệnh do các loài virus, vi khuẩn khác nhau. Các bước thực hiện PCR phức tương tự như kỹ thuật PCR chuẩn, yêu cầu tính toán tỉ lệ các cặp mồi và lượng dNTP, enzym DNA polymerase thích hợp để hiệu quả PCR chính xác.



**Hình 5.9. Sơ đồ kỹ thuật PCR phức (Multiplex PCR)**

#### 4.7. Kỹ thuật PCR tại chỗ (PCR in situ)

Kỹ thuật PCR tại chỗ (PCR in situ) là phương pháp PCR để nhận các gen, các đoạn DNA nào đó ngay trên mẫu tế bào hoặc mô không cần tách chiết và tinh sạch DNA. PCR in situ được sử dụng để phát hiện khả năng bị nhiễm các loại virus khác nhau ở người và động vật: phát hiện Human Papilovirus, HIV (Nuovo và cộng sự, 1992); HCV (Komminoth và cộng sự, 1994); HBV (Long và cộng sự, 1993). Kỹ thuật PCR in situ có nhiều tên gọi khác nhau: in cell PCR, PCR in situ hybridization...

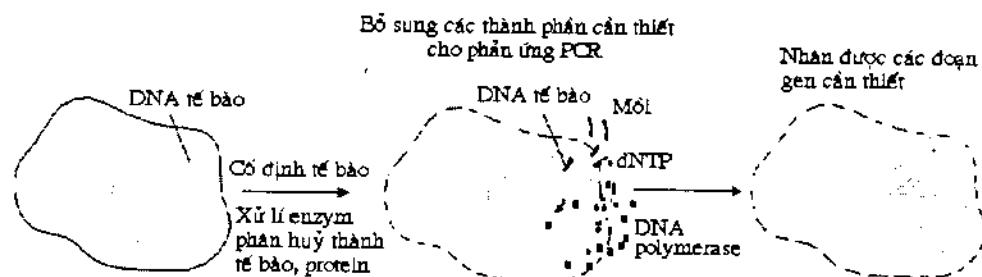
Kỹ thuật PCR tại chỗ gồm một số bước cơ bản sau:

Cố định tế bào và mô của các mẫu nghiên cứu trên lam kính, theo các phương pháp thông thường của phòng thí nghiệm.

Xử lý enzym lysozym phân hủy thành tế bào, sau đó xử lý enzym proteinase phân hủy protein của tế bào.

Bổ sung các thành phần cần thiết cho phản ứng PCR như thực hiện kỹ thuật PCR chuẩn, riêng các loại dNTP thường sử dụng loại dNTP đánh dấu huỳnh quang.

Đặt các lam kính vào các thiết bị PCR chuyên dụng, để thực hiện kỹ thuật nhân gen. Kiểm tra kết quả PCR dưới kính hiển vi huỳnh quang.



Hình 5. 10. Sơ đồ kỹ thuật PCR tại chỗ (*in situ PCR*)

Trên đây là một số kỹ thuật PCR thường được sử dụng trong các nghiên cứu kỹ thuật gen và công nghệ sinh học phân tử. Ngoài ra, còn có rất nhiều kỹ thuật PCR khác như PCR định lượng (Real-time PCR), PCR phát hiện đột biến, Short PCR, Long PCR... ngày càng có độ chính xác cao hơn, tiện dụng hơn sử dụng trong các nghiên cứu sinh học.

## II. Phương pháp điện di

### 1. Nguyên lý chung

Phương pháp điện di là kỹ thuật thí nghiệm được sử dụng phổ biến trong nghiên cứu cấu trúc và đặc điểm sinh học của các phân tử mang điện tích (DNA, protein và enzym...). DNA là đại phân tử sinh học mang điện tích âm, do đó trong môi trường có điện trường, các phân tử DNA có kích thước khác nhau di chuyển từ cực âm sang cực dương với tốc độ khác nhau.

Thiết bị điện di gồm nguồn điện, buồng điện di, hệ đệm thích hợp và các loại bảm gel khác nhau (agarose, polyacrylamid, tinh bột...) được sử dụng nhiều

trong nghiên cứu độ tinh sạch DNA, kích thước các đoạn DNA cắt bởi enzym giới hạn hoặc kiểm tra kết quả tách dòng gen...

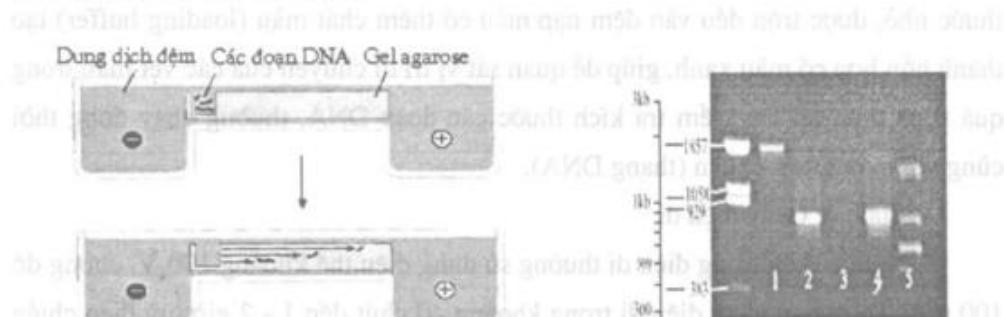
### 2. Các kỹ thuật điện di chủ yếu

Tùy theo mục đích nghiên cứu và loại bàn gel sử dụng, người ta chia thành các phương pháp điện di khác nhau, thường được gọi theo tên loại bàn gel sử dụng. Các kỹ thuật điện di chủ yếu gồm:

Điện di trên gel agarose: là kỹ thuật điện di để kiểm tra DNA sử dụng agarose từ 0,3 % - 2,0% làm bàn gel, tùy theo kích thước của DNA. Nếu kích thước đoạn DNA nhỏ dưới 1 kb, hàm lượng agarose sử dụng khoảng 1 - 2 %, trong các phòng thí nghiệm thường sử dụng hàm lượng agarose khoảng 0,7 % đến 0,8%.

Điện di trên gel polyacrylamid: là kỹ thuật điện di để kiểm tra DNA hoặc protein, sử dụng bàn polyacrylamid với nồng độ khoảng 3,5% - 20% làm bàn gel, tùy theo kích thước đoạn DNA hoặc trọng lượng phân tử protein. Đoạn gen có kích thước càng nhỏ, protein có khối lượng phân tử nhỏ thì hàm lượng polyacrylamid sử dụng càng lớn, và ngược lại.

Điện di trên giấy: là kỹ thuật điện di sử dụng giấy thấm Whatman 3 MM làm cầu nối giữa 2 điện cực của buồng điện di, các đại phân tử tích điện dịch chuyển trên giấy với tốc độ khác nhau tùy theo kích thước và trọng lượng phân tử. Trước đây, phương pháp điện di trên giấy thường sử dụng trong nghiên cứu đa hình di truyền protein, izoenzym...



Hình 5.11. Sơ đồ buồng điện di nằm ngang và kết quả điện di

Điện di trên gel tinh bột: tinh bột biến tính được sử dụng làm bàn gel dùng trong điện di trong các nghiên cứu protein, enzym và izoenzym...

Ngoài ra, tuỳ thuộc thiết kế vị trí các cực của buồng điện di, có thể chia làm hai nhóm phương pháp là điện di nằm ngang và điện di đứng. Phương pháp điện di nằm ngang có buồng điện di với các cực điện di trên cùng mặt phẳng, còn phương pháp điện di đứng với buồng điện di có cực âm (-) thường ở phía trên, cực dương (+) ở phía dưới.

### *3. Các bước thực hiện kỹ thuật điện di*

#### Bước 1: Chuẩn bị dung dịch đệm và bàn gel

Dung dịch đệm thường sử dụng trong điện di agarose và polyacrylamid là TBE (Tris - Boric acid- EDTA) và TAE (Tris - Acetic acid- EDTA).

Bản gel được chuẩn bị bằng dung dịch đệm 1x TBE (1x TAE) hoà tan với hàm lượng agarose (polyacrylamid) với hàm lượng cần thiết, đặt trong lò vi sóng hoặc dun nóng đến khi tan hoàn toàn, đổ vào khuôn có các lược cài sẵn để tạo bản gel với số giếng nắp mẫu cần thiết.

Khi bản gel đã đông cứng, đặt bản gel vào buồng điện di và đổ dung dịch đệm ngập bản gel khoảng 1-2 mm.

#### Bước 2: Nạp mẫu

Mẫu DNA đã tách chiết và xử lí enzym giới hạn cắt thành các đoạn kích thước nhỏ, được trộn đều vào đệm nạp mẫu có thêm chất màu (loading buffer) tạo thành hỗn hợp có màu xanh, giúp dễ quan sát vị trí di chuyển của các vệt mẫu trong quá trình điện di. Để kiểm tra kích thước các đoạn DNA, thường chạy đồng thời cùng với mẫu DNA chuẩn (thang DNA).

#### Bước 3: Chạy điện di

Nguồn điện trong điện di thường sử dụng điện thế khoảng 100 V, cường độ 100 mA. Thời gian chạy điện di trong khoảng 40 phút đến 1 - 2 giờ tuỳ theo chiều dài bản gel và mức độ yêu cầu tách các phân tử (có thể dựa vào vị trí các vạch màu, khi các vạch màu chạy được khoảng 3/4 chiều dài bản gel có thể kết thúc điện di).

**Bước 4: Nhuộm bản gel và soi gel**

Kết thúc điện di, bản gel được lấy ra nhuộm ethyldium bromid (EtBr), hoặc các phương pháp nhuộm khác (nhuộm bạc, nhuộm SYBR green...).

Bản gel sau khi nhuộm EtBr được đặt vào thiết bị hiện hình có tia UV, để quan sát kết quả điện di. Vị trí các đoạn điện di theo kích thước khác nhau quan sát được những vệt sáng màu da cam trên bản gel, hoặc đặt bản gel vào các thiết bị chụp ảnh chuyên dụng để chụp ảnh..., trường hợp nhuộm bạc hoặc nhuộm SYBR green có thể xác định kết quả bằng mắt thường.

### **III. Phương pháp giải trình tự gen**

Giải trình tự gen (DNA sequencing) là phương pháp xác định vị trí xấp xep các nucleotid trong phân tử DNA. Giải trình tự bộ gen của người, động vật và vi sinh vật giúp chẩn đoán bệnh tật và nhiều ứng dụng trong thực tiễn sản xuất, phục vụ lợi ích con người. Có 3 nhóm phương pháp giải trình tự gen chủ yếu: phương pháp Maxam và Gilbert, phương pháp dideoxy (phương pháp Sanger) và phương pháp giải trình tự gen bằng máy tự động.

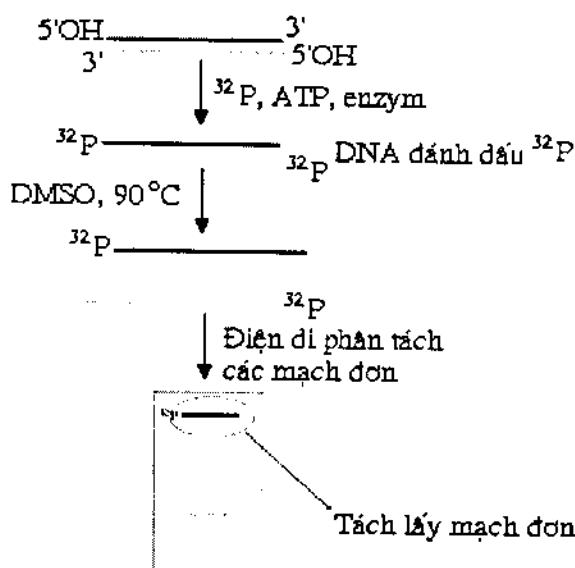
#### ***1. Giải trình tự DNA theo phương pháp Maxam và Gilbert***

##### **1. 1. Nguyên lý**

Năm 1977 Alan Maxam và Walter Gilbert phát minh phương pháp giải trình tự gen theo phương pháp hoá học. Tạo mạch khuôn DNA trên cơ sở đánh dấu đồng vị phóng xạ P<sup>32</sup> đầu 5' và biến tính phân tử DNA thành các mạch đơn không tự xoắn lại với nhau. Các mạch đơn đã đánh dấu phóng xạ được phân vào các ống nghiệm khác nhau, thực hiện kỹ thuật xử lí hoá học đặc hiệu để phân cắt các mạch đơn thành các đoạn ngắn hơn kém một nucleotid, từ đó xác định trình tự DNA bằng phương pháp điện di.

### 1. 2. Các bước thực hiện

Bước 1: Đánh dấu phóng xạ  $P^{32}$  ở đầu 5' của mạch khuôn, biến tính và điện di tách lấy một mạch DNA làm mạch khuôn, cho các xử lí tiếp theo.



**Hình 5.12. Sơ đồ đánh dấu phóng xạ và biến tính thu DNA mạch đơn**

Bước 2: Thực hiện các phản ứng hoá học đặc hiệu

Có thể sử dụng 4 - 5 ống nghiệm để thực hiện các phản ứng hoá học đặc hiệu khác nhau.

Ống	Xử lí đặc hiệu	Vị trí cắt
1	Dimethyl sunfat	Cắt tại G
2	Acid pH = 2	Cắt tại A và G
3	Hydrazin	Cắt tại C và T
4	Hydrazin + NaCl	Cắt tại C
5	NaOH	Cắt tại A hoặc C (A>C)

Trong mỗi ống nghiệm có hàm lượng nhất định các mạch đơn DNA khuôn dã đánh dấu phóng xạ, được xử lí hoá học đặc hiệu khác nhau, tạo các đoạn DNA cắt dài ngắn khác nhau.

Bước 3: Lấy sản phẩm xử lí trong mỗi ống nghiệm đem chạy điện di, hiện hình phóng xạ và xác định kết quả.

Ví dụ, mạch đơn DNA sử dụng làm khuôn có trình tự các nucleotid:

$^{32}\text{P}$ - ATGCCAGTTGTACAGC

- Xử lí dimethyl sunfat ở ống phản ứng 1, phản ứng cắt tại G tạo thành các đoạn mạch đơn như sau:

$^{32}\text{P}$ -ATC

$^{32}\text{P}$ -ATGCCA

$^{32}\text{P}$ -ATGCCAGTT

$^{32}\text{P}$ -ATGCCAGTTGTACCA

- Xử lí acid pH = 2 ở ống phản ứng 2 cắt tại vị trí A và G:

Cắt tại A tạo thành các đoạn mạch đơn :

$^{32}\text{P}$ -ATGCC

$^{32}\text{P}$ -ATGCCAGTTGT

$^{32}\text{P}$ -ATGCCAGTTGTACCA

Cắt tại G tạo thành các đoạn mạch đơn :

$^{32}\text{P}$ -ATC

$^{32}\text{P}$ -ATGCCA

$^{32}\text{P}$ -ATGCCAGTT

$^{32}\text{P}$ -ATGCCAGTTGTACCA

Bước 4: Tập hợp kết quả thu được ở tất cả các ống nghiệm, thu được các đoạn polynucleotid dài ngắn hơn kém 1 nucleotid, từ đó xác định được trình tự của mạch DNA khuôn.

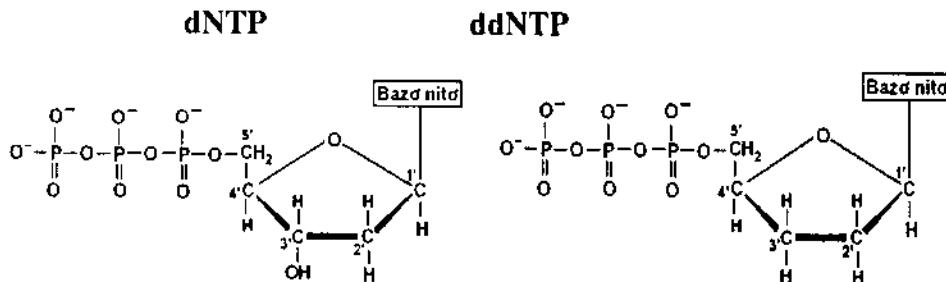
Phương pháp xác định trình tự nucleotid bằng cách phản ứng hoá học đặc hiệu của Maxam và Gilbert có độ chuẩn xác kém, cần phải thực hiện nhiều lần và loại bỏ các sai sót để chọn một kết quả gần đúng nhất. Do đó, phương pháp Maxam và Gilbert ít được ứng dụng trong thực tiễn.

## 2. Giải trình tự DNA theo phương pháp dideoxy

### 2. 1. Nguyên lý

Xác định trình tự gen theo phương pháp dideoxy do F. Sanger, Smith và Coulson phát hiện năm 1977, dựa trên cơ chế tổng hợp DNA trong cơ thể sinh vật. F. Sanger đã giải trình tự hoàn chỉnh bộ gen của thực khuẩn thĕ φX 174 kí sinh vi khuẩn *E. coli* (1977), là bộ gen đầu tiên được xác định trình tự.

Chuẩn bị các điều kiện cần thiết đảm bảo thực hiện tổng hợp DNA (Mạch khuôn DNA đánh dấu phóng xạ, dNTP tự do các loại, enzym DNA polymerase, mồi...). Trong quá trình tổng hợp mạch đơn DNA, bổ sung một hàm lượng nhỏ dideoxynucleotid (ddNTP) mỗi loại. Do ddNTP bị mất cả 2 nhóm OH (ở carbon số 2 và số 3), nên trong quá trình tổng hợp mạch đơn DNA, gặp ngẫu nhiên một dideoxynucleotid làm cho phản ứng tổng hợp DNA ngừng lại. Kết quả tổng hợp DNA tạo nên các chuỗi mạch đơn mới hơn kém nhau một nucleotid, từ đó xác định được trình tự của mạch khuôn DNA.



Hình 5.13. Sơ đồ cấu trúc dNTP và ddNTP

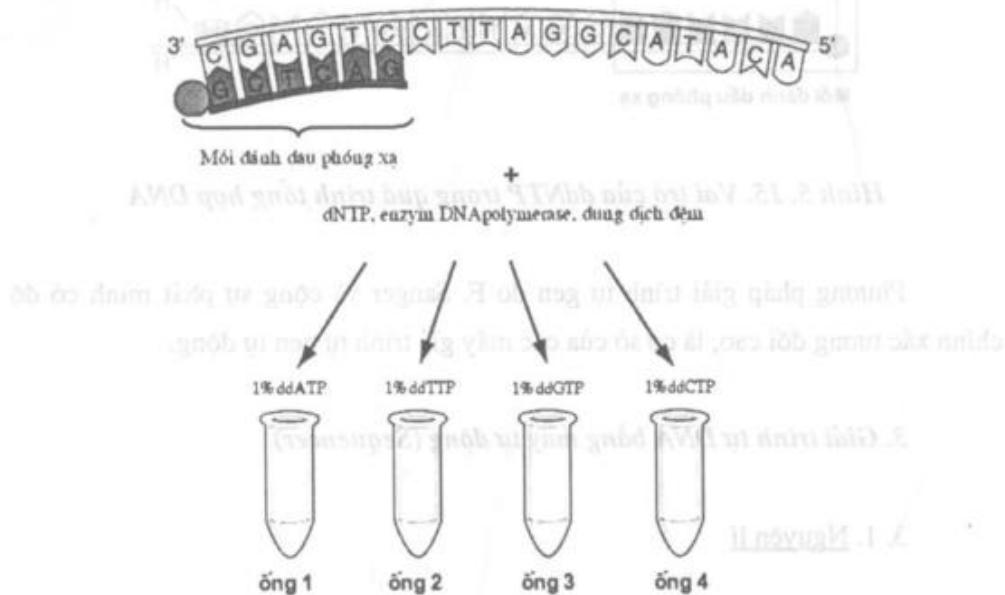
### 2. 2. Các bước thực hiện:

Bước 1: Biến tính DNA khuôn, điện di trên gel polyacrylamid (có bổ sung ure) để thu các mạch đơn DNA.

Bước 2: Chuẩn bị phản ứng tổng hợp DNA

Lấy bốn ống Eppendorf, cho các thành phần cần thiết vào mỗi ống (mạch khuôn, mỗi có dánh dấu phỏng xạ, enzym DNA polymerase và dNTP (dATP, dTTP, dGTP, dCTP) và dung dịch đậm thích hợp). Thêm vào mỗi ống tương ứng một lượng nhỏ (khoảng 1%) một loại ddNTP nhất định: ống 1 bổ sung 1% ddATP, ống 2 thêm 1% ddTTP, ống 3 thêm 1% ddGTP và ống 4 thêm 1% ddCTP (hình 5.14).

Bước 3: Thực hiện phản ứng tổng hợp DNA với các thiết bị tuân hoàn nhiệt.

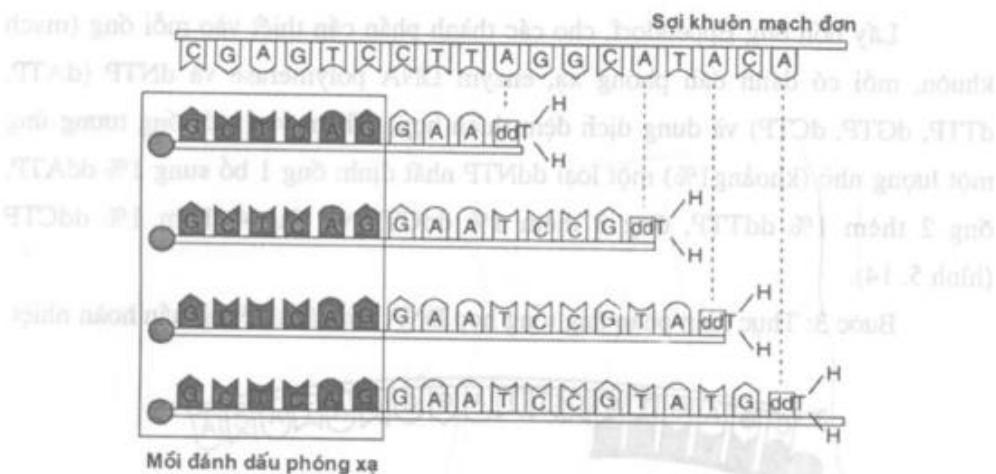


**Hình 5.14. Sơ đồ chuẩn bị tổng hợp các mạch đơn DNA**

Bước 4: Điện di kết quả, so sánh các kết quả các chuỗi mạch đơn DNA được tổng hợp để xác định trình tự của mạch khuôn.

Ví dụ, ống 2 có bổ sung 1% ddTTP, nên khi mạch đơn DNA đang tổng hợp gặp ngẫu nhiên một ddT, phản ứng tổng hợp mạch đơn DNA mới lập tức ngừng lại.

Tổng hợp kết quả ở các 4 ống, thu được các đoạn mạch đơn hơn kẽm nhau một nucleotid, thu được trình tự đoạn mạch đơn mới được tổng hợp và xác định được trình tự của mạch DNA khuôn.



Hình 5.15. Vai trò của ddNTP trong quá trình tổng hợp DNA

Phương pháp giải trình tự gen do F. Sanger và cộng sự phát minh có độ chính xác tương đối cao, là cơ sở của các máy giải trình tự gen tự động.

### 3. Giải trình tự DNA bằng máy tự động (Sequencer)

#### 3.1. Nguyên lý

Máy giải trình tự gen bán tự động hoặc tự động hoàn toàn được thiết kế trên nguyên tắc sử dụng ddNTP do F. Sanger và cộng sự phát minh.

Trong quá trình tổng hợp DNA có sử dụng các mồi và dNTP đánh dấu huỳnh quang thay cho đánh dấu phóng xạ, mỗi loại dNTP được đánh dấu huỳnh quang khác nhau, biểu thị các màu sắc khác nhau.

Máy thực hiện tổng hợp các mạch đơn DNA mới trên cả 2 mạch khuôn DNA, đồng thời sử dụng các phần mềm tính toán trên máy tính để xử lý kết quả giúp kiểm soát được các sai sót, đảm bảo độ chính xác cao.

#### 3.2. Các bước thực hiện

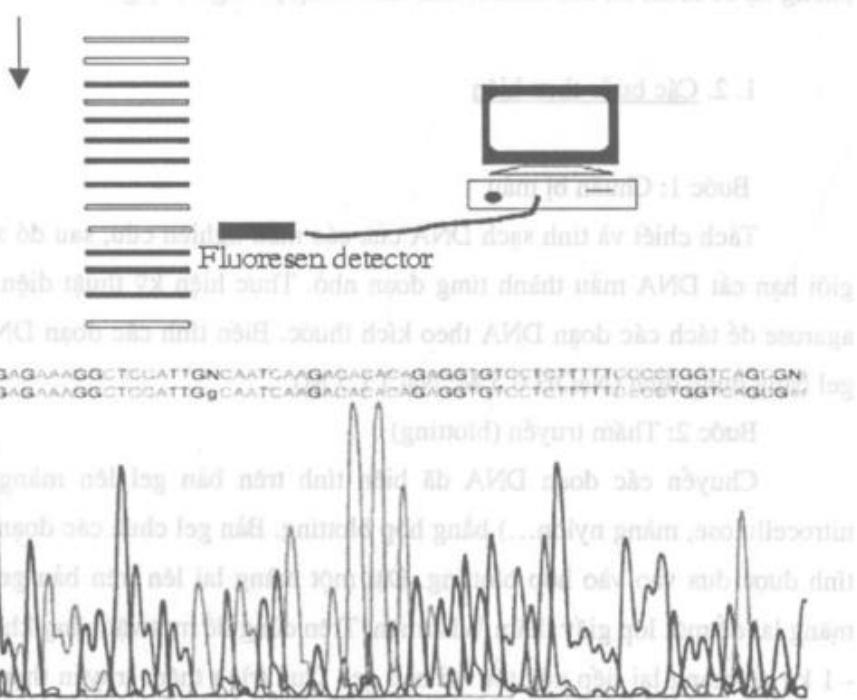
Bước 1: Chuẩn bị DNA khuôn

DNA khuôn phải đảm bảo độ tinh sạch cao, có kích thước thích hợp (thường sử dụng các mẫu DNA đã được tách dòng, để có độ chính xác cao).

Bước 2: Chuẩn bị các điều kiện cần thiết cho phản ứng tổng hợp DNA (theo protocol của các hãng cung cấp hóa chất, hoặc thiết bị...) và thực hiện nhân DNA bằng thiết bị PCR.

Bước 3: Nạp sản phẩm PCR vào máy giải trình tự gen (sequencer) và chạy máy giải trình tự gen. Trong quá trình máy hoạt động các nucleotid được kiểm soát nhờ detecter báo về bộ phận xử lý của máy tính, phần mềm xử lý kết quả và vẽ đồ thị của các loại nucleotid đi qua detecter.

Bước 4: Phân tích kết quả từ các dữ liệu do máy tính cung cấp, so sánh kết quả xác định trình tự ở cả 2 mạch đơn mới, đồng thời căn cứ vào đồ thị để sửa các sai sót và chọn kết quả đúng nhất.



Hình 5.16. Sơ đồ kỹ thuật giải trình tự DNA bằng máy tự động

## IV. Các phương pháp lai phân tử

### 1. Kỹ thuật lai Southern blotting

Kỹ thuật lai Southern blotting do Edward Southern phát minh 1975, là phương lai giữa các đoạn DNA mạch đơn của mẫu cần nghiên cứu, với các mẫu dò đánh dấu phóng xạ (các đoạn oligo nucleotid đã biết trình tự) nhằm kiểm tra sự có mặt của các trình tự đích.

#### 1. 1. Nguyên lí

Các đoạn DNA được biến tính tách rời các mạch đơn và được chuyển lên màng lai. Bằng phản ứng lai các đoạn mạch đơn trên màng lai với mẫu dò đánh dấu phóng xạ để kiểm tra các trình tự cần thiết bằng phóng xạ tự ghi.

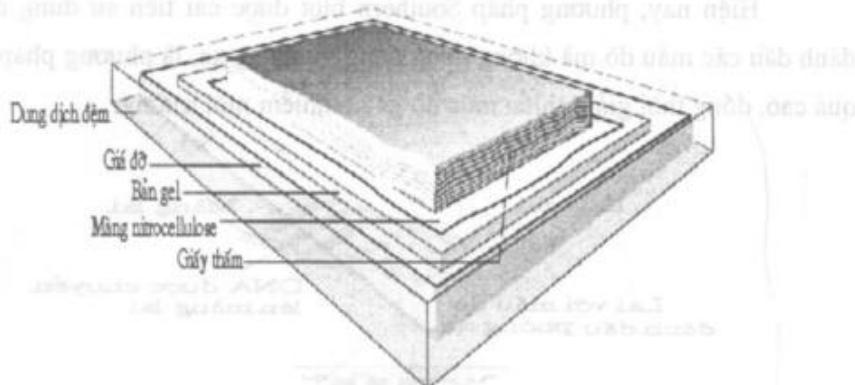
#### 1. 2. Các bước thực hiện

##### Bước 1: Chuẩn bị mẫu

Tách chiết và tinh sạch DNA của các mẫu nghiên cứu, sau đó xử lí enzym giới hạn cắt DNA mẫu thành từng đoạn nhỏ. Thực hiện kỹ thuật điện di trên gel agarose để tách các đoạn DNA theo kích thước. Biến tính các đoạn DNA trên bàn gel bằng dung dịch (NaOH 0,5 M; NaCl 1,5 M).

##### Bước 2: Thấm truyền (blotting)

Chuyển các đoạn DNA đã biến tính trên bàn gel lên màng lai (màng nitrocellulose, màng nylon...) bằng hộp blotting. Bàn gel chứa các đoạn DNA biến tính được đưa vào vào hộp blotting. Đặt một màng lai lên trên bàn gel, phía trên màng lai để một lớp giấy thấm Whatman. Trên cùng để một vật nặng khoảng 0,5 kg - 1 kg để màng lai tiếp xúc tốt với bàn gel. Quá trình thấm truyền thực hiện trong thời gian 8 – 12 giờ (qua đêm), để đảm bảo các đoạn mạch đơn DNA được đưa lên màng lai.



**Hình 5. 17. Sơ đồ hộp thám truyền (blotting)**

#### Bước 3: Lai phân tử

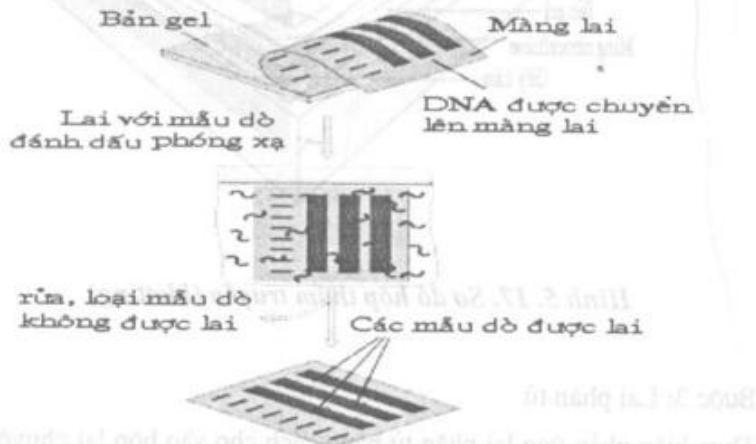
Thực hiện phản ứng lai phân tử bằng cách cho vào hộp lai chuyên dụng một lớp khoảng 0,2 - 0,4 cm dịch lai, bổ sung một hàm lượng cần thiết các mẫu dò đặc hiệu đánh dấu phóng xạ. Lấy màng lai sau khi kết thúc thám truyền đưa vào hộp lai để chế độ nhiệt  $65^{\circ}\text{C}$  trong 3 - 8 giờ, sau đó thêm dịch lai với chế độ nhiệt  $65^{\circ}\text{C}$ , lắc nhẹ để tăng tốc độ phản ứng. Rửa màng lai ở nhiệt độ  $65^{\circ}\text{C}$  bằng dung dịch SSC (NaCl 3M, natri citrat 0,3 M, và nước, pH = 7,0) từ hai đến ba lần để loại bỏ hết các mẫu dò không được lai. Thám khô màng lai nhẹ nhàng bằng giấy lọc hoặc được sấy khô ở  $65^{\circ}\text{C}$ .

#### Bước 4: Xác định kết quả bằng kỹ thuật phóng xạ tự ghi

Màng lai sau khi đã làm khô, lấy một tấm phim nhạy X quang đặt lên màng lai và để trong một thiết bị chuyên dụng ở nhiệt độ  $-20^{\circ}\text{C}$ . Sau đó xử lý dưới ánh sáng tử ngoại trong 4 giây để kiểm tra kết quả lai.

Phương pháp Southern blot được ứng dụng rộng rãi trong công nghệ sinh học phân tử, giúp phát hiện nhanh một trình tự DNA nào đó với các mẫu dò đã biết trước trình tự, cho phép kiểm tra các bệnh do ung thư, các bệnh do nhiễm vius và vi khuẩn ở người và động thực vật.

Hiện nay, phương pháp Southern blot được cải tiến sử dụng digoxigenin đánh dấu các mẫu dò mà không dùng đồng vị phóng xạ, là phương pháp lai có hiệu quả cao, đồng thời giảm thiểu mức độ gây ô nhiễm môi trường.



**Hình 5.18. Sơ đồ kỹ thuật lai Southern blotting**

## 2. Kỹ thuật lai Northern blotting

Kỹ thuật Northern blotting là kỹ thuật do Edward Southern cải tiến, nhằm phát hiện các trình tự nhất định của phân tử RNA, bằng cách lai RNA với các mẫu dò DNA tạo nên các phân tử lai RNA-DNA.

Nguyên tắc thực hiện kỹ thuật Northern blotting gồm các bước cơ bản giống như kỹ thuật lai Southern blot. Mẫu tế bào được tách chiết RNA và tinh sạch, sau đó RNA được chuyển lên màng lai bằng kỹ thuật thẩm truyền. Thực hiện phản ứng lai với các mẫu dò DNA đánh dấu phóng xạ, tạo nên các phân tử lai RNA - DNA. Phát hiện các phân tử lai bằng kỹ thuật phóng xạ tia X quang phay phóng xạ.

### 3. Kỹ thuật lai Western blotting

Kỹ thuật lai Western blotting là kỹ thuật lai protein với kháng thể đặc hiệu, sử dụng trong các nghiên cứu protein và enzym. Kỹ thuật lai Western blotting gồm một số bước cơ bản sau:

- Chọn mẫu và thực hiện kỹ thuật tách chiết protein với các dung dịch đậm thích hợp.
- Diện di protein trên gel polyacrylamid.
- Thực hiện kỹ thuật thấm truyền (blotting) để chuyển protein lên màng lai.
- Lai protein trên màng lai với các protein kháng thể đặc hiệu.
- Hiện hình, nhuộm màu và kiểm tra kết quả.

### 4. Kỹ thuật lai tại chỗ (*in situ hybridization*)

Kỹ thuật lai tại chỗ là phương pháp lai phân tử, ngay trong tế bào không cần tách chiết acid nucleic. Lai tại chỗ thường sử dụng mẫu dò được đánh dấu bằng huỳnh quang thay cho mẫu dò đánh dấu phóng xạ. Tuỳ từng loại tế bào và mô có thể thực hiện các kỹ thuật lai tại chỗ khác nhau.

#### 4. 1. Kỹ thuật lai trên nhiễm sắc thể

Kỹ thuật lai trên nhiễm sắc thể có thể thực hiện với các bước chủ yếu sau:

- Làm tiêu bản nhiễm sắc thể: chọn tế bào ở kỳ giữa của quá trình phân bào khi các nhiễm sắc thể có kích thước lớn nhất. Sử dụng các kỹ thuật xử lí và cố định tế bào trên lam kính, để định vị hình dạng và vị trí của nhiễm sắc thể. Xử lí loại bỏ RNA và protein bằng enzym RNase và protease.

- Thực hiện phản ứng lai: sử dụng hỗn dịch chứa các mẫu dò DNA (hoặc RNA) đã đánh dấu huỳnh quang (có thể dùng mẫu dò đánh dấu đồng vị phóng xạ, tuỳ theo mục đích nghiên cứu) nhỏ lên vị trí mẫu đã cố định trên lam kính, để thực hiện phản ứng lai.

- Xác định kết quả: rửa mẫu lai bằng các dung dịch rửa thích hợp, và quan sát kết quả trên kính hiển vi, ở những vị trí có các phân tử lai xuất hiện các chấm có màu (tùy theo loại chất huỳnh quang đánh dấu mẫu dò).

Căn cứ vào loại mẫu dò đã biết trình tự, có thể xác định được vị trí nhiễm sắc thể có các gen cần nghiên cứu.

#### 4. 2. Kỹ thuật lai trên tế bào và mô

Kỹ thuật lai trên tế bào và mô thường sử dụng để phát hiện hoặc kiểm tra sự hoạt động của một gen nào đó. Trong mỗi cơ thể, các gen khác nhau hoạt động trong các mô, tế bào ở những giai đoạn khác nhau. Hoạt động phiên mã của gen tạo nên các loại mRNA tương ứng. Kỹ thuật lai trên tế bào và mô nhằm nghiên cứu giai đoạn hoạt động của gen, mối liên quan giữa phiên mã và giải mã...

Kỹ thuật lai trên tế bào và mô gồm một số bước cơ bản sau:

- Cố định tế bào và mô: bằng các kỹ thuật tế bào học phù hợp (tẩm parafin, khử nước...), cắt lát siêu mỏng ( $7 - 10 \mu\text{m}$ ) và cố định trên lam kính.

- Xử lý enzym protease để loại bỏ protein, sau đó xử lý bằng enzym DNase để loại bỏ DNA.

- Lai với mẫu dò DNA đặc hiệu đã đánh dấu phóng xạ hoặc huỳnh quang.

- Rửa mẫu để loại các mẫu dò không được lai và xác định kết quả dưới kính hiển vi chuyên dụng (kính hiển vi huỳnh quang hoặc phản pha).

### V. Một số kỹ thuật phân tử sử dụng trong xác định tính đa dạng di truyền của sinh vật và ứng dụng trong phân loại phân tử

#### I. Kỹ thuật RAPD

Kỹ thuật RAPD (Random Amplified Polymorphism DNA - đa hình các đoạn DNA được khuếch đại ngẫu nhiên) do William phát minh 1990, Welsh và cộng sự hoàn thiện 1991.

### 1. 1. Nguyên lí

Sử dụng cùng một số cặp mồi ngẫu nhiên nhất định (thường sử dụng từ 10 - 40 cặp mồi) để thực hiện phản ứng PCR, nhằm nhận các đoạn DNA đặc trưng của các mẫu nghiên cứu. Nếu các mẫu nghiên cứu có bộ gen giống nhau hoàn toàn, sản phẩm PCR thu được gồm các đoạn DNA hoàn toàn giống nhau về kích thước và cấu trúc. Khi bộ gen của các mẫu nghiên cứu có sự khác biệt nhau, kết quả PCR nhận được các đoạn khác biệt nhau.

Mồi ngẫu nhiên là các đoạn oligo nucleotid gồm khoảng 8 - 20 nucleotid, đặc trưng với mỗi loài sinh vật.

### 1. 2. Các bước thực hiện

- Tách chiết và tinh sạch DNA bộ gen của các mẫu nghiên cứu, và thực hiện phản ứng PCR nhận các đoạn DNA với các cặp mồi đã lựa chọn (trong cùng điều kiện giống nhau).

- Xác định tính hệ số đồng dạng di truyền hoặc mức độ khác nhau giữa các mẫu nghiên cứu bằng các phần mềm phân tích sinh học thông dụng (NTSYS - PC, UPGMA Cluster, Gelcompar, PHILIP...) hoặc tính theo các biểu thức khác nhau bởi các nghiên cứu của các nhà khoa học.

Có thể tính hệ số đồng dạng di truyền theo biểu thức của M. Nei và Li (1979):

$$S_y = \frac{2N_{ij}}{N_i + N_j}$$

trong đó  $N_i$  là số vạch của giống i,  $N_j$  là số vạch của giống j và  $N_{ij}$  là số vạch chung của cả giống i và giống j.

Kỹ thuật RAPD là kỹ thuật phân loại phân tử dễ sử dụng, thực hiện nhanh được ứng dụng trong xác định tính đa dạng sinh học và quan hệ họ hàng của các giống thực vật, động vật khác nhau trong các loài. Ví dụ, sử dụng 43 cặp mồi ngẫu nhiên do hãng Operon Technology sản xuất, Henry Nguyễn (Đại học Texas - Mỹ) đã phát hiện sự khác biệt di truyền của 13 giống lúa có nguồn gốc khác nhau ở châu Á, châu Âu và trung Á.

Hạn chế chủ yếu của kỹ thuật RAPD là độ chuẩn xác không cao, kết quả phân tích không ổn định. Chẳng hạn, cùng sử dụng số loại cặp mồi ngẫu nhiên và điều kiện PCR giống nhau, nhưng ở các lần phân tích khác nhau cho kết quả không giống nhau. Do đó, trong phân tích di truyền để đảm bảo kết quả có độ chính xác cao, thường sử dụng kết hợp kỹ thuật RAPD với các kỹ thuật phân tử khác.

## **2. Kỹ thuật RFLP**

RFLP (Restriction Fragment Length Polymorphism DNA - đa hình chiều dài các đoạn DNA cắt bởi các enzym giới hạn). Kỹ thuật RFLP dựa trên đặc điểm của các loại enzym giới hạn khác nhau, tạo nên các đoạn cắt DNA khác nhau phân biệt được bằng điện di đồ, các đoạn cắt còn được gọi là các “dấu vân tay” đặc trưng cho từng phân tử DNA. Bản đồ di truyền các kết quả RFLP, có tính chính xác cao thường sử dụng trong nghiên cứu sự khác biệt trong cấu trúc bộ gen của các cá thể, các loài sinh vật, nhằm so sánh sự khác biệt giữa các mẫu nghiên cứu, xác định nguồn gốc hoặc kiểm tra mức độ tiến hóa của các loài sinh vật...

### **2. 1. Nguyên lý**

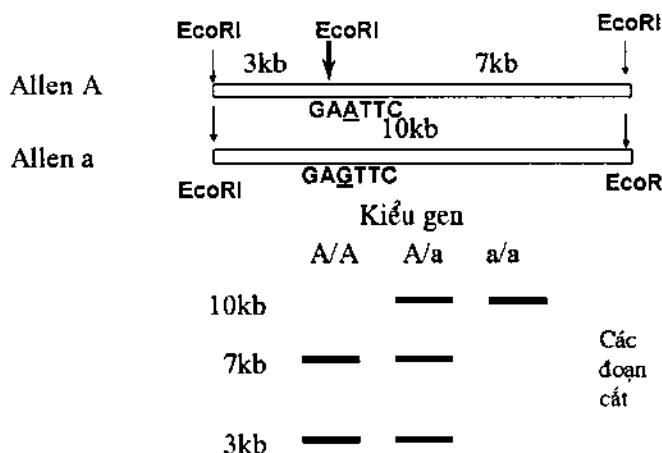
Các mẫu nghiên cứu được tách chiết và tinh sạch DNA, sau đó xử lý các mẫu DNA bằng cùng một số loại enzym giới hạn. Mỗi enzym giới hạn nhận biết và cắt đặc hiệu DNA ở những vị trí xác định. Do đó, các bộ gen có cấu trúc khác nhau tạo nên số lượng đoạn cắt DNA khác nhau, kích thước của các đoạn cắt có thể khác nhau.

Ngược lại, những bộ gen hoàn toàn giống nhau tạo nên số lượng, kích thước các đoạn cắt DNA giống nhau, có thể được phát hiện trên điện di đồ.

## 2. 2. Các bước thực hiện

Kỹ thuật RFLP còn được gọi là kỹ thuật Fingerprinting. Qui trình thực hiện RFLP gồm các bước cơ bản:

- Tách chiết và tinh sạch DNA của các mẫu nghiên cứu.
- Xử lý các mẫu DNA trong cùng điều kiện với cùng một số loại enzym giới hạn thích hợp.
- Điện di và xác định đa dạng di truyền các đoạn cắt giới hạn, lập bản đồ di truyền các đoạn giới hạn.



Hình 5.19. Sơ đồ kỹ thuật RFLP

Trong chọn giống thực vật, chọn lọc các cá thể mang các gen lặn mong muốn có thể thực hiện nhanh bằng sử dụng kỹ thuật RFLP, không cần phân tích qua lai nhiều thế hệ. Kỹ thuật RFLP có thể được sử dụng phát hiện các cặp gen

nghiên cứu là đồng hợp tử trội hay di hợp tử, hoặc có sự liên kết giữa các gen...

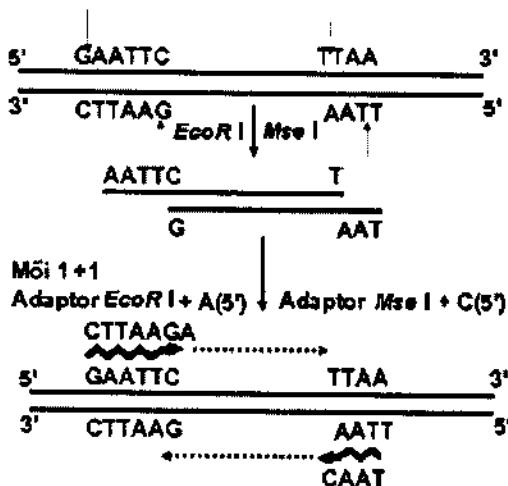
Kỹ thuật RFLP được sử dụng để kiểm tra sự phân li di truyền của một số tính trạng theo qui luật Mendel, hoặc ứng dụng trong chọn động vật, chọn giống thực vật hoặc so sánh sự khác nhau giữa các cá thể, các loài sinh vật...

### 3. Kỹ thuật AFLP

Kỹ thuật AFLP (Amplified Fragment Length Polymorphism - đa hình chiều dài các đoạn DNA được khuếch đại chọn lọc) do Vos và cộng sự phát minh 1975.

#### 3. 1. Nguyên lý

Kỹ thuật AFLP dựa trên cơ sở sử dụng PCR với các cặp mồi đặc hiệu để khuếch đại một đoạn DNA đặc trưng của các mẫu nghiên cứu. Sau đó thực hiện PCR lần thứ hai với các mồi được nối thêm 1 hoặc 2 nucleotid ở đầu 3' (hoặc 5') của mồi, để nhận các đoạn DNA có tính đặc hiệu cao. Kết quả PCR cho phép phân biệt mức độ giống nhau hoặc khác nhau giữa các mẫu nghiên cứu ngay cả các mẫu so sánh có quan hệ di truyền gần nhau (mẫu DNA của các đơn vị dưới loài).



Hình 5. 20. Sơ đồ kỹ thuật AFLP

### 3. 2. Các bước thực hiện

- Tách chiết và tinh sạch DNA của các mẫu nghiên cứu và xử lí cắt DNA bằng các cặp enzym giới hạn thích hợp, bổ sung các adaptor đặc hiệu tạo các đoạn có đầu mút giống nhau, đặc trưng cho các mồi đặc hiệu đã lựa chọn.
- Sử dụng kỹ thuật PCR nhân đặc hiệu các đoạn DNA với các cặp mồi có bổ sung 1 hoặc 2 nucleotid.
- Phân tích sản phẩm PCR trên điện di đồ và so sánh kết quả của các mẫu nghiên cứu bằng các phần mềm thông dụng.

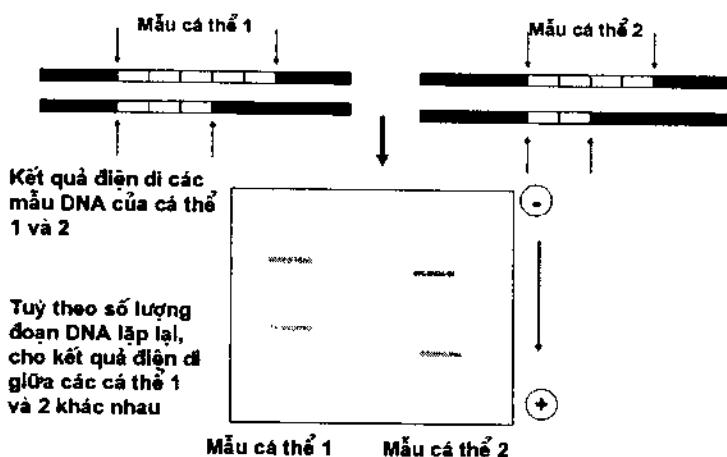
### 4. Kỹ thuật VNTR

Kỹ thuật vi vê tinh (VNTR-Variable Number Tandem Repeats) là kỹ thuật kiểm tra các đoạn lặp lại ngắn trong bộ gen, còn được gọi là kỹ thuật SSR (simple sequence repeat - các trình tự lặp đơn giản).

#### 4. 1. Nguyên lý

Bộ gen sinh vật bậc cao có nhiều đoạn DNA lặp lại, các đoạn lặp DNA microsatellite có kích thước ngắn từ 2 – 6 nucleotid, có số lần lặp lại đặc trưng với mỗi loài, mỗi giống. Ví dụ, ở lúa có khoảng gần 1000 lần lặp lại trật tự AC/TG, khoảng trên 300 lần lặp lại trật tự GATA/CTAT, nhiều loài cây một lá mầm (ngô, lúa...) lặp lại đoạn CGG/GCC.

Sử dụng kỹ thuật PCR với các cặp mồi đơn giản, có thể khuếch đại được đoạn DNA lặp lại của các mẫu nghiên cứu. Dựa vào thang DNA chuẩn và độ dài mỗi đoạn lặp DNA, có thể tính được số lượng đoạn lặp DNA trong bộ gen, hoặc trong mỗi gen. Số lượng đoạn lặp DNA trong gen, hình thành các alell khác nhau, tạo nên dãy di truyền đa alell hay mức đa dạng di truyền của từng locus gen, của các mẫu nghiên cứu.

**Hình 5. 21. Sơ đồ kỹ thuật VNTR (SSR)**

#### 4. 2. Các bước thực hiện

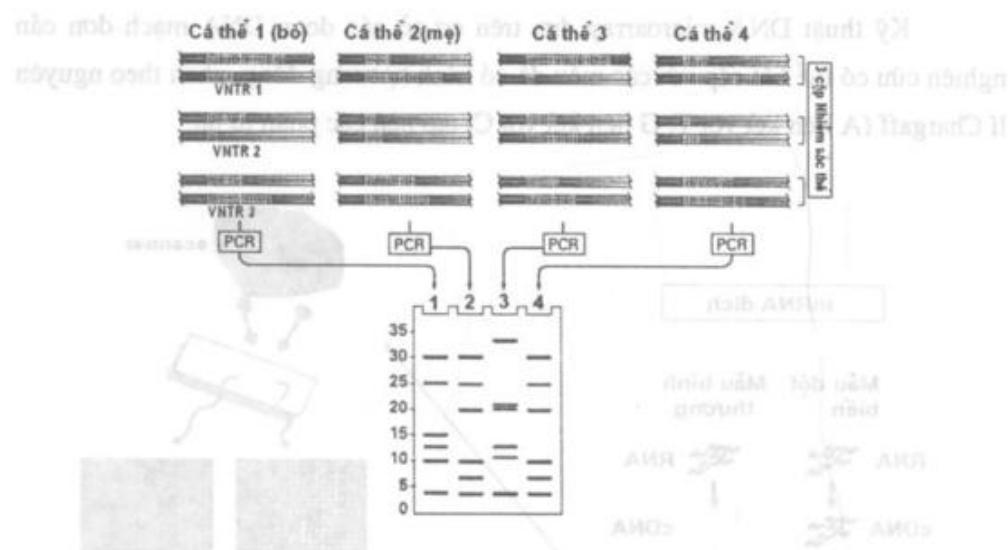
Kỹ thuật VNTR hay SSR gồm một số bước chủ yếu:

- Tách chiết và tinh sạch DNA của các mẫu nghiên cứu, thực hiện phản ứng PCR với các cặp mồi đặc trưng cho các đoạn lặp đơn giản để khuếch đại các đoạn DNA lặp lại trong gen hoặc bộ gen.

- Điện di kết quả PCR trên gel agarose với thang DNA chuẩn, đồng thời dựa vào kích thước đoạn lặp để xác định số lượng đoạn lặp và độ đa dạng di truyền của các mẫu.

- Xử lí số liệu, tính toán kết quả để xác định mức độ giống và khác nhau giữa các mẫu nghiên cứu. Có thể sử dụng một số phần mềm chuyên dụng (PHILIP, NTSYSpc...) xây dựng cây phân loài...

Ví dụ, bằng kỹ thuật VNTR (SSR) nghiên cứu sự khác biệt bộ gen giữa cây đu đủ đực và đu đủ cái, Parasnis (1998) phát hiện đoạn lặp DNA kích thước 5 kb lặp lại rất nhiều lần một trình tự GATA, có ở cây đực mà không có ở cây cái. Một số nghiên cứu sử dụng kỹ thuật VNTR phát hiện trong bộ gen cây cà chua có 32 đoạn lặp khác nhau giữa các giống.



*Hình 5. 22. Sơ đồ kỹ thuật VNTR trong so sánh quan hệ di truyền giữa các cá thể sinh vật*

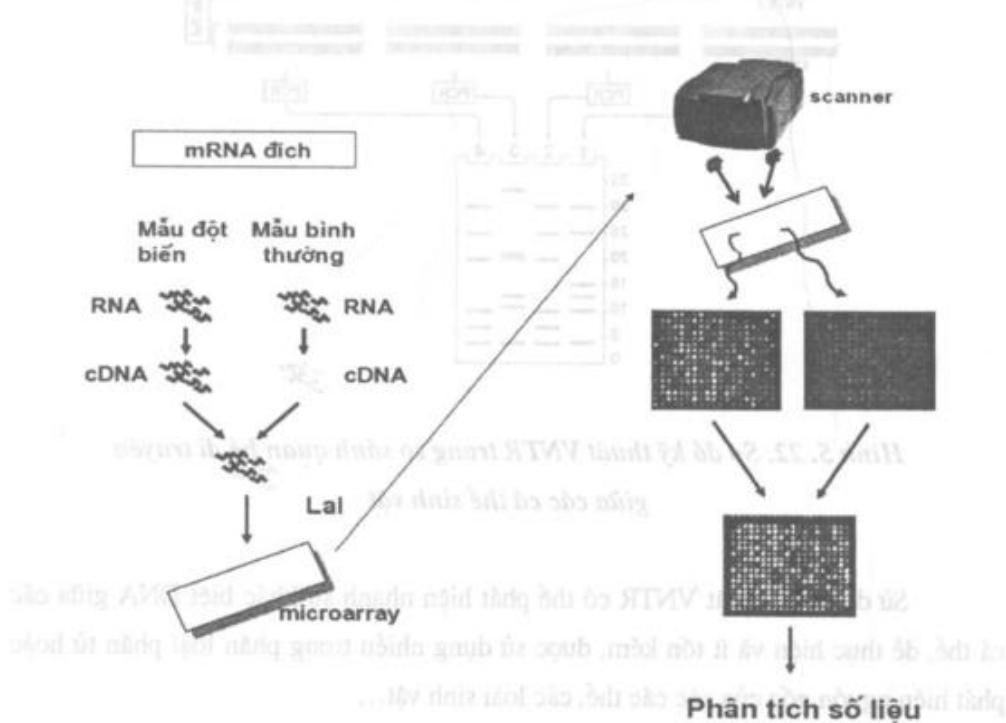
Sử dụng kỹ thuật VNTR có thể phát hiện nhanh sự khác biệt DNA giữa các cá thể, dễ thực hiện và ít tốn kém, được sử dụng nhiều trong phân loại phân tử hoặc phát hiện nguồn gốc của các cá thể, các loài sinh vật...

### 5. Kỹ thuật DNA microarray

#### 5. 1. Nguyễn lí

Kỹ thuật DNA microarray là kỹ thuật lai các đoạn cDNA mạch đơn ngắn được đánh dấu huỳnh quang, với tập hợp (mảng - array) các đoạn oligonucleotid đã biết trình tự được gắn ở các vị trí xác định trên GeneChip, để kiểm tra và phát hiện một trình tự nucleotid nào đó, phát hiện các đột biến, nghiên cứu cấu trúc gen... Mỗi GeneChip chứa từ 1000 000 - 20 000 000 mẫu dò là các oligonucleotid đã biết trình tự.

Kỹ thuật DNA microarray dựa trên cơ sở các đoạn DNA mạch đơn cần nghiên cứu có thể bắt cặp với các mẫu dò có trình tự tương đồng, nhau theo nguyên lý Chargaff (A liên kết với T, G liên kết với C) tạo nên các phân tử lai.



Hình 5. 22. Nguyên lý chung của kỹ thuật microarray

Dựa vào kỹ thuật phát hiện huỳnh quang, với các máy quét scanner chuyên dụng có thể xác định được vị trí của các phân tử lai, từ đó xác định được trình tự của đoạn DNA mạch đơn, hoặc trình tự gen.

### 5. 2. Các bước thực hiện

Bước 1: Thu thập các mẫu sinh học cần nghiên cứu, phá vỡ tế bào và tách chiết RNA thông tin

Bước 2: Phiên mã ngược

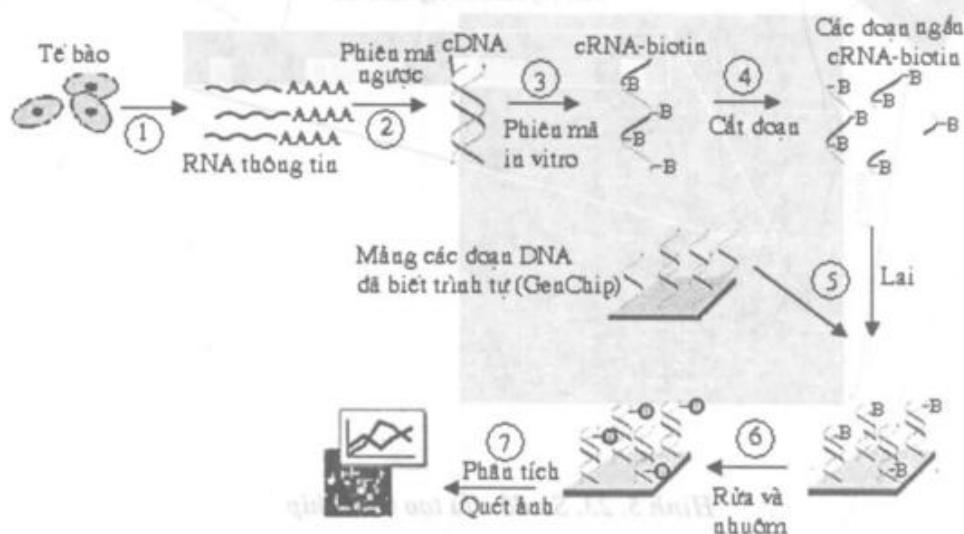
Thực hiện kỹ thuật phiên mã ngược với các mồi oligo dT, enzym phiên mã ngược và dNTP các loại tạo nên các phân tử cDNA tương ứng với các mRNA.

Bước 3: Phiên mã in vitro tạo các mạch cRNA đánh dấu huỳnh quang

Sử dụng T7 RNA polymerase, biotin - UTP và biotin - CTP, từ mỗi mạch đơn cDNA tạo nên khoảng 50 - 100 cRNA đánh dấu Biotin.

Bước 4: Cắt đoạn

Ü cRNA – biotin ở 94°C trong dung dịch đậm thích hợp, cắt thành các đoạn cRNA-biotin khoảng 35 – 200 ribonucleotid.



**Hình 5. 23. Sơ đồ các bước thực hiện kỹ thuật microarray**

Bước 5: Lai

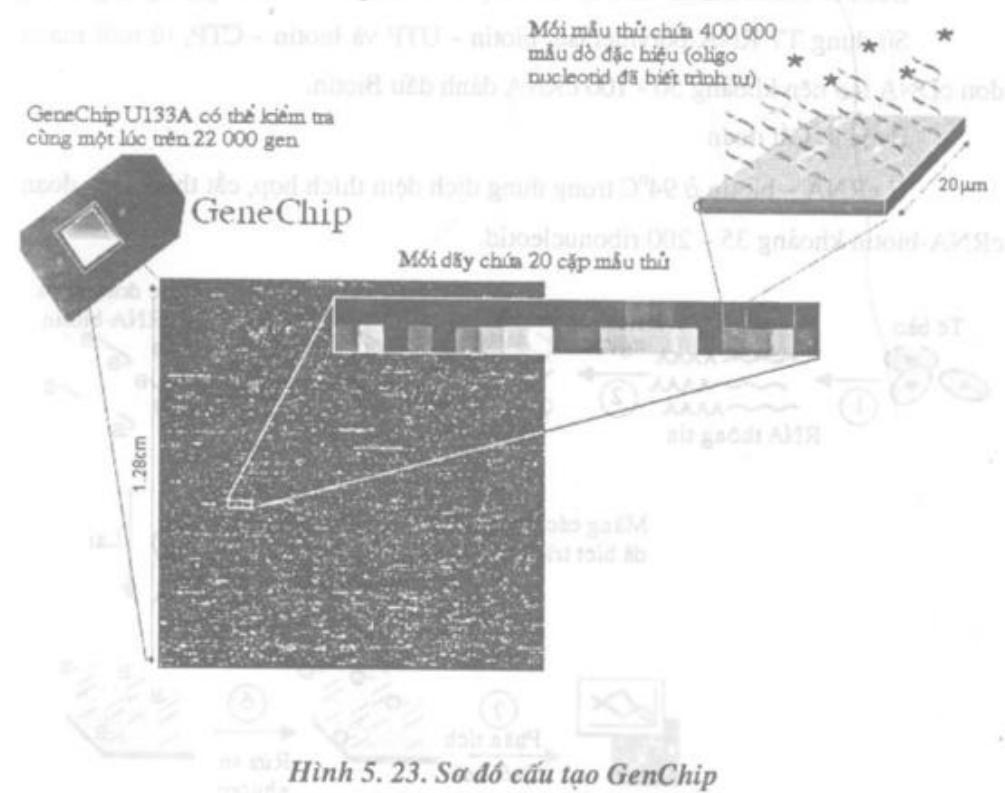
Các đoạn cRNA-biotin khoảng 35 – 200 ribonucleotid được chuyển lên các GeneChip chứa các mẫu dò cDNA mạch đơn đã biết trình tự, thực hiện phản ứng lai.

Bước 6: Nhuộm

Rửa mảng lai (GeneChip) để loại bỏ các cRNA đánh dấu huỳnh quang không được lai, sau đó nhuộm các phân tử lai bằng streptavidin-phycoerythrin.

### Bước 7: Hiện hình và phân tích

Các GeneChip đã nhuộm được đưa vào các thiết bị scaner chuyên dụng quét dữ liệu và phân tích kết quả.



Hình 5.23. Sơ đồ cấu tạo GenChip

Hiện nay, kỹ thuật microarray là một trong những kỹ thuật phân tử được ứng dụng trong chẩn đoán có độ chính xác rất cao các bệnh ung thư, bệnh do rối loạn di truyền, bệnh nhiễm trùng và các đột biến gen... ở người và vật nuôi.

## VI. Các phương pháp chuyển gen

Chuyển gen là tập hợp nhiều kỹ thuật khác nhau, nhằm đưa một gen lạ (một đoạn DNA) vào trong tế bào, gắn gen đó với bộ gen của tế bào làm cho gen lạ tồn tại lâu dài và ổn định và được biểu hiện, tạo nên các thế chuyển gen hay sinh vật biến đổi gen (GMO - Genetically Modified Organism).

Các kỹ thuật chuyển gen được chia làm hai nhóm phương pháp: chuyển gen trực tiếp và chuyển gen gián tiếp.

- Nhóm phương pháp chuyển gen trực tiếp hay còn gọi là nhóm phương pháp chuyển gen vật lí gồm các kỹ thuật chủ yếu sau:

Biến nạp DNA trực tiếp (Direct DNA uptake): siêu âm, PEG (Polyethylen glycol)...

Kỹ thuật điện xung (Electroporation)

Kỹ thuật vi tiêm (Microinjection)

Kỹ thuật bắn gen (Biolistic hay gene gun)

Kỹ thuật chuyển gen qua ống phun...

- Nhóm các phương pháp chuyển gen gián tiếp gồm các kỹ thuật chuyển gen nhờ các nhân tố trung gian, chủ yếu nhờ các vectơ sinh học là vi khuẩn *Agrobacterium tumefaciens* hoặc chuyển gen nhờ vectơ virus.

### *1. Chuyển gen bằng thiết bị siêu âm*

#### *1.1. Nguyễn lý*

Siêu âm có khả năng tạo nên các lỗ màng, làm cho gen cần chuyển (các đoạn DNA) dễ dàng vào trong tế bào, có thể tạo được các thể tái tổ hợp với bộ gen của tế bào chủ.

Dựa vào các gen chỉ thị, thực hiện quá trình sàng lọc để lựa chọn các thể chuyển gen (GMO - Genetically Modified Organism). Nuôi cấy hoặc chăm sóc các thể chuyển gen và kiểm tra sự biểu hiện gen mới trong thể chuyển gen, tiếp tục chọn lọc để tạo được các thể chuyển gen theo ý muốn có tính ổn định di truyền.

#### *1.2. Các bước thực hiện*

Kỹ thuật chuyển gen bằng siêu âm thường sử dụng trong chuyển gen thực vật, gồm một số bước cơ bản:

- Lựa chọn đối tượng chuyển gen, chọn gen cần chuyển và vectơ mang gen cần chuyển (vectơ chuyển gen).
- Lấy mẫu (lá non, mỏ vò cây...), thực hiện kỹ thuật tạo tế bào trần (protoplast) bằng cách xử lý với enzym pectinase và cellulase, li tâm thu tế bào trần. Tế bào trần được treo trong môi trường chứa 20% - 21% đường sucrose, mật độ tế bào khoảng  $3 \times 10^6$  / ml.
- Thực hiện kỹ thuật biến nạp bằng thiết bị siêu âm với tần số 20 kHz trong thời gian theo từng chu kỳ ngắn khoảng 10 phần nghìn giây, tổng thời gian khoảng 0,5 - 0,8 giây.
- Nuôi tế bào ở môi trường thích hợp để phục hồi thành tế bào, chọn lọc các tế bào được chuyển gen nhờ các gen chỉ thị.
- Kiểm tra hoạt tính của gen đã chuyển trong thể chuyển gen (GMO)

## ***2. Chuyển gen bằng kỹ thuật điện xung (Electroporation)***

### **2.1. Nguyên lí**

Kỹ thuật điện xung là một trong các kỹ thuật biến nạp có hiệu quả cao, được ứng dụng trong chuyển gen với tế bào thực vật và động vật. Tạo các tế bào trần (protoplasts) và sử dụng thiết bị điện xung (Electroporation) với điện thế cao khoảng 500 V/cm trong thời gian 4 -5 phần nghìn giây, tạo nên các lỗ trên màng tế bào trần, làm cho DNA lật bên ngoài có thể xâm nhập vào bộ gen của tế bào.

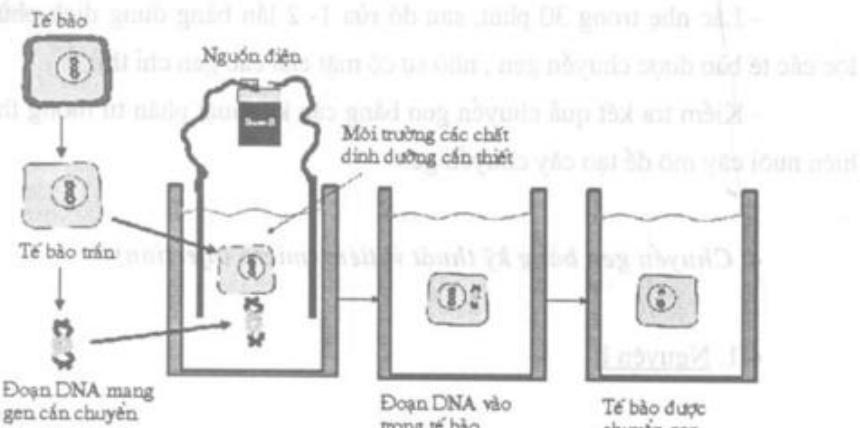
Quá trình tái tổ hợp giữa gen lạ với bộ gen làm cho gen lạ được gắn vào bộ gen tế bào, hình thành các thể chuyển gen.

### **2.2. Các bước thực hiện**

- Tạo các tế bào khả biến (đối với thực vật thường tạo các tế bào trần để tăng hiệu quả chuyển gen), tạo dịch huyền phù tế bào khả biến với vectơ tái tổ hợp mang gen cần chuyển.

Thực hiện kỹ thuật điện xung với cường độ  $0,25 \mu\text{F}$ ,  $200 \Omega$ ,  $1,5 \text{ kV}$  trong  $4 - 5 / 1000$  giây. Các lỗ màng tạo nên do xung điện cho phép DNA lọt dễ vào trong tế bào.

Sàng lọc các tế bào đã được chuyển gen nhờ các gen chỉ thị. Nuôi cấy tế bào, kiểm tra hoạt tính của gen mới ở thể chuyển gen (GMO).



Hình 5. 24. Sơ đồ kỹ thuật chuyển gen bằng xung điện

### 3. Chuyển gen bằng kỹ thuật PEG (Polyethylen glycol)

#### 3.1. Nguyên lý

Polyethylen glycol (PEG) là chất có ái lực lớn với nước. Khi ở nồng độ cao, PEG làm cho DNA dính vào màng nguyên sinh chất của tế bào, tế bào trán đưa PEG có DNA bám quanh vào tế bào theo cơ chế ẩm bào.

Trong tế bào DNA được giải phóng, PEG được đưa ra khỏi tế bào. DNA lọt có thể tái tổ hợp với bộ gen tế bào ở những điểm tương đồng, tạo nên các tế bào được chuyển gen.

#### 3.2. Các bước thực hiện

- Chuyển gen bằng PEG thường được thực hiện có hiệu quả với các tế bào trán ở thực vật, do đó cần tạo các tế bào trán bằng enzym pectinase và cellulose. Tế bào trán được treo lơ lửng trong môi trường dinh dưỡng phù hợp.

- Li tâm làm lắng các tế bào trán và bổ sung 20 µl vectơ chứa gen cần chuyển (20µg DNA) plasmid tái tổ hợp có gen cần chuyển và thêm 2ml PEG.

- Lắc nhẹ trong 30 phút, sau đó rửa 1- 2 lần bằng dung dịch phù hợp, sàng lọc các tế bào được chuyển gen , nhờ sự có mặt của các gen chỉ thị.

- Kiểm tra kết quả chuyển gen bằng các kỹ thuật phân tử thông thường, thực hiện nuôi cấy mô để tạo cây chuyển gen.

#### **4. Chuyển gen bằng kỹ thuật vi tiêm (micro injection)**

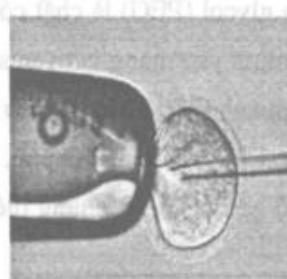
##### **4.1. Nguyên lí**



Sử dụng thiết bị vi tiêm đưa gen cần chuyển vào trong tế bào chủ ở các vị trí cần thiết. Tạo điều kiện tái tổ hợp của gen cần chuyển với bộ gen của tế bào, sau đó lựa chọn các thể chuyển gen.

##### **4.2. Các bước thực hiện**

Kỹ thuật vi tiêm thường được sử dụng để chuyển gen, chuyển nhân lưỡng bội vào tế bào trứng rỗng để tạo động vật chuyển gen.

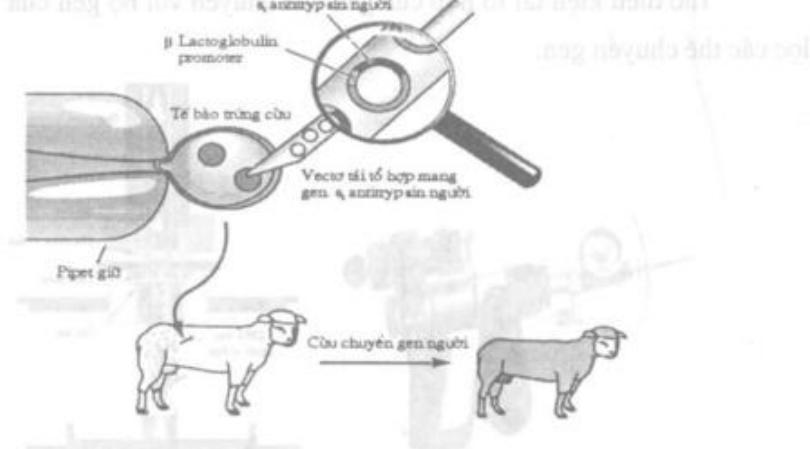


Hình 5. 25. Thiết bị vi tiêm và kỹ thuật vi tiêm vào tế bào trứng

- Lựa chọn tế bào đích là các tế bào trứng đơn bội (n), thực hiện hút bỏ nhân tạo các trứng rỗng nhờ thiết bị vi tiêm.

- Hút lấy nhân của tế bào cho (có thể là nhân của tế bào soma hoặc nhân của hợp tử mới thụ tinh...) và đưa vào trong trứng rỗng. Kích thích xung điện với tần số thích hợp để phôi phát triển.

- Chuyển phôi vào tử cung của một cái "mẹ nuôi", để tạo động vật chuyển gen, và kiểm tra hoạt động của gen được chuyển ở động vật chuyển gen.



Hình 5. 26. Sơ đồ chuyển gen  $\alpha_1$  antitrypsin người vào cừu bằng vi tiêm

### 5. Chuyển gen bằng súng bắn gen

Kỹ thuật chuyển gen bằng súng bắn gen - "Gen gun" đầu tiên do Đại học Cornell (Mỹ) đề xuất 1987, với súng bắn gen PDS -1000/He (particle delivery system 1000/He, hãng Biorad sản xuất). Ngày nay, có nhiều loại súng bắn gen khác nhau gồm các loại súng bắn gen nhỏ và súng bắn gen hệ thống, có hiệu quả khác nhau trong chuyển gen.

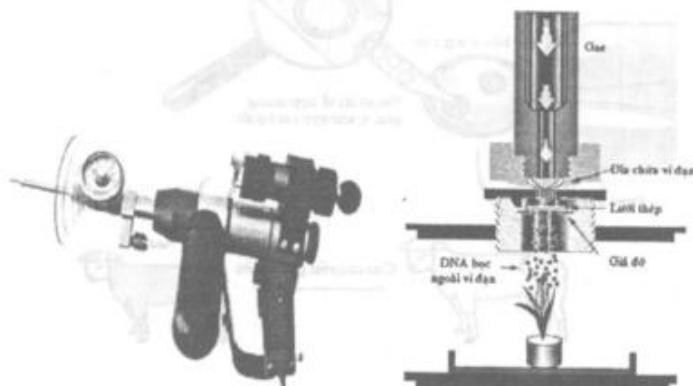
#### 5.1. Nguyễn lý

Gen cần chuyển được trộn với phụ gia, tạo thành một lớp bao bên ngoài vi-

đạn. Vi đạn (micro projectile) dùng bằng vàng nguyên chất hoặc tungsten kích thước 0,5 - 1 µm. Vi đạn trộn với vectơ chuyển gen được làm khô trên một đĩa kim loại mỏng đường kính 0,5 - 0,8 cm.

Súng bắn gen có áp suất khí heli đưa vi đạn mang các gen cần chuyển vào trong tế bào chủ tới các vị trí cần thiết, đĩa kim loại được giữ lại nhờ lưỡi thép ở đầu của súng. Một số loại súng bắn gen không sử dụng đĩa kim loại, khí nén mạnh có thể đẩy trực tiếp vi đạn (súng bắn gen Sautter 1991 hoặc súng Finer 1992)...

Tạo điều kiện tái tổ hợp của gen cần chuyển với bộ gen của tế bào và chọn lọc các thế chuyển gen.



A. Súng bắn gen loại nhỏ B. Hệ thống bắn gen PDS-1000/He

Hình 5. 27. Sơ đồ súng bắn gen

### 5.2. Các bước thực hiện

Căn cứ vào đặc điểm của loại mô và tế bào đích đã lựa chọn để chọn kích thước vi đạn, khoảng cách từ đầu nòng súng đến mô, tế bào.

- Lựa chọn loại vectơ mang gen cần chuyển và phụ gia thích hợp, tạo một lớp màng chứa gen cần chuyển bao quang vi đạn.

Cần căn cứ vào loại tế bào đích để lựa chọn kích thước vi đạn, loại vi đạn cũng như áp suất khí heli thích hợp để đảm bảo hiệu quả chuyển gen. Trong thực tế, sử dụng các thiết bị bắn gen khác nhau cần chú ý các thông số của hệ thống để đảm

bảo hiệu quả cao. Ví dụ, hệ thống súng bắn gen PDS-1000/He có thể sử dụng các thông số ở bảng sau:

Loại tế bào cần chuyển gen	Khoảng cách đầu nòng đến đích	Áp suất khí heli (psi)	Loại vi đạn
Ví khuẩn	6 cm	1100	0,4 µm (tungsten)
Nấm men	6 cm	1300	0,6 µm (vàng)
Ví tảo	6 cm	1300	0,6 µm (vàng)
Phôi (thực vật)	6 cm	1300	1,0 µm (vàng)
Tế bào thực vật	9 cm	1100	1,0 µm (vàng)
Tế bào động vật	3 cm	1100	1,6 µm (vàng)

- Thực hiện kỹ thuật bắn gen và sàng lọc các tế bào đã được chuyển gen, nuôi cấy mô và kiểm tra hoạt động của gen mới ở thể chuyển gen (GMO).

Thử nghiệm chuyển gen bằng súng bắn gen Corbo-PIG, một số nghiên cứu ở Việt Nam đã chuyển các gen chống rầy, chống nấm bạc lá cho lúa, bước đầu thu được một số dòng lúa chuyển gen có khả năng chống bệnh tương đối tốt.

## 6. Kỹ thuật chuyển gen bằng súng nhiệt

### 6.1. Nguyễn lý

Sử dụng hiệu ứng chênh lệch nhiệt độ, làm thay đổi tính thẩm chọn lọc của màng tế bào vi khuẩn, cho phép DNA lạ dễ dàng xâm nhập tế bào.

Hiện tượng tái tổ hợp giữa gen lạ với bộ gen của tế bào chủ, tạo các tế bào mang các gen theo mong muốn.

### 6.2. Các bước thực hiện

Chuyển gen bằng sốc nhiệt thực hiện có hiệu quả khi chuyển gen ở vi sinh vật và tế bào trứng sau thụ tinh ở một số loài cá.

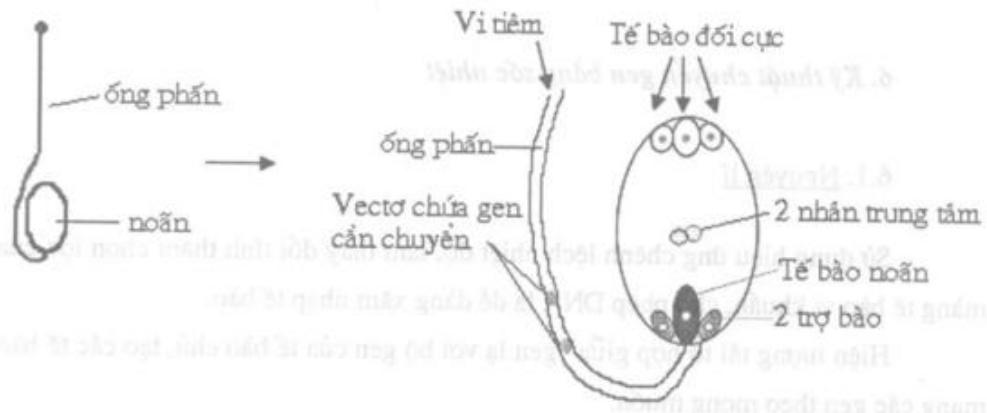
- Chuyển gen bằng sốc nhiệt cần tạo các tế bào khả biến.
- Thực hiện kỹ thuật biến nạp bằng sốc nhiệt theo qui trình thông thường trong phòng thí nghiệm (chương 4).
- Sàng lọc các tế bào được chuyển gen, thử hoạt tính của gen mới.

### 7. Kỹ thuật chuyển gen qua ống phẩn

Kỹ thuật chuyển gen qua ống phẩn đã tạo nên nhiều giống thực vật chuyển gen (lúa và một số loại cây ăn quả) có năng suất cao chất lượng tốt.

Kỹ thuật chuyển gen qua ống phẩn được thực hiện theo một số bước cơ bản sau:

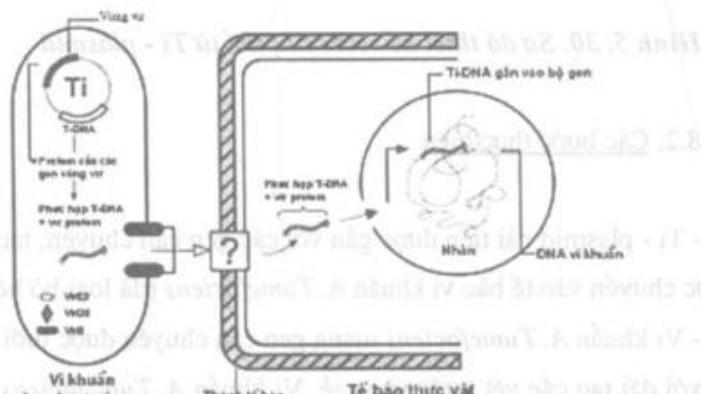
- Cắt ngang ống dẫn phẩn, và thực hiện kỹ thuật đưa dung dịch chứa vectơ tái tổ hợp mang gen cần chuyển tới noãn.
- Kiểm tra hoạt động của gen đã chuyển, chọn lọc thể chuyển gen.
- Sau đó thực hiện giao phẩn chéo một vài thế hệ, kết hợp với chọn lọc tạo các giống chuyển gen có tính ổn định di truyền cao.



Hình 5.28. Sơ đồ kỹ thuật chuyển gen qua ống phẩn

#### 8. Kỹ thuật chuyển gen nhờ vi khuẩn *Agrobacterium tumefaciens*

Vi khuẩn *A. tumefaciens* có khả năng xâm nhiễm tự nhiên cây trồng, qua các vết xước ở rễ cây chuyển Ti - plasmid vào cây gây các khối u ở nhiều loài thực vật. Ti - plasmid (Tumour inducing), kích thước khoảng 140 - 200 kb. Vi khuẩn *A. rhizogenes* xâm nhiễm rễ cây qua vết xước, chuyển Ri - plasmid (Root inducing) vào cây, gây hiện tượng rễ phu hoặc rễ tơ ở cây.



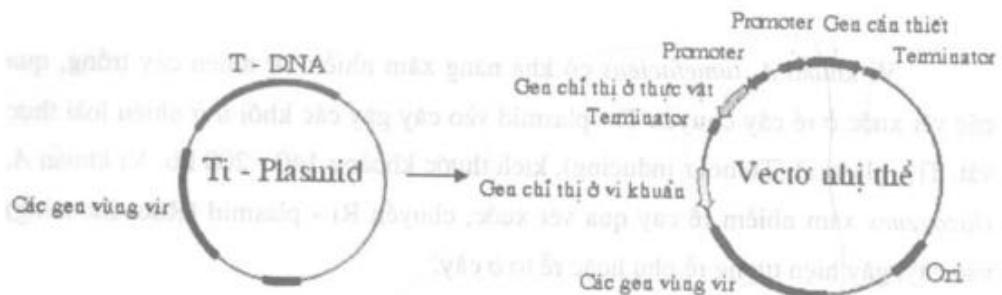
Hình 5.29. Cơ chế xâm nhiễm của Ti-plasmid

Plasmid của *A. tumefaciens* mang đoạn T - DNA gồm các gen mã hoá tổng hợp auxin, opin, xitokin gây hiện tượng các tế bào phân chia không kiểm soát, tạo các khối u ở cây.

### 8.1. Nguyên lí

Ti - plasmid của vi khuẩn *A. tumefaciens* được cải biến cấu trúc di truyền bằng cách cắt bỏ đoạn một đoạn dài chứa T- DNA, đồng thời gắn thêm một đoạn

của plasmid khác (pBR322...), tạo nên các vectơ chuyển gen có kích thước phù hợp, có các gen giúp khả năng xâm nhiễm tự nhiên của vi khuẩn vào thực vật.

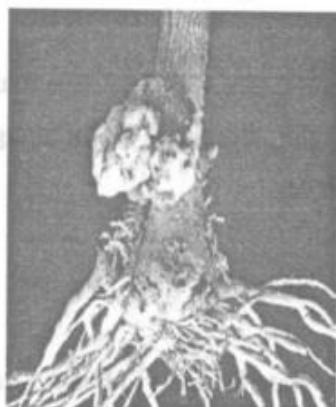


Hình 5. 30. Sơ đồ thiết kế vectơ nhị thể từ Ti - plasmid

## 8.2. Các bước thực hiện

- Ti - plasmid cài tiến được gắn với các gen cần chuyển, tạo các vectơ tái tổ hợp, được chuyển vào tế bào vi khuẩn *A. Tumefaciens* (đã loại bỏ hết Ti - plasmid).

- Vi khuẩn *A. Tumefaciens* mang gen cần chuyển được tưới vào vùng rễ cây kết hợp xới đất tạo các vết xước vùng rễ. Vi khuẩn *A. Tumefaciens* xâm nhiễm cây qua vết xước ở rễ, đưa gen cần chuyển vào cây. Kiểm tra và sàng lọc để thu được các cây chuyển gen.



Hình 5. 31. Khối u ở thực vật do Ti-plasmid gây nên

- Có thể chuyển gen nhờ vi khuẩn *A. Tumefaciens* mang vectơ tái tổ hợp bằng cách trộn các mẫu lá nhỏ với vi khuẩn vi khuẩn *A. Tumefaciens* trong hộp Petri có môi trường dinh dưỡng phù hợp. Sau đó chuyển các mẫu lá sang môi trường tạo callus, dựa vào gen chỉ thị để lựa chọn các callus được chuyển gen. Thực hiện kỹ thuật nuôi cây mô để tạo cây chuyển gen.

Ngoài các kỹ thuật chuyển gen chủ yếu trên, còn nhiều kỹ thuật chuyển gen được sử dụng tạo sinh vật chuyển gen như chuyển gen nhờ virus, chuyển gen nhờ kỹ thuật điện di...

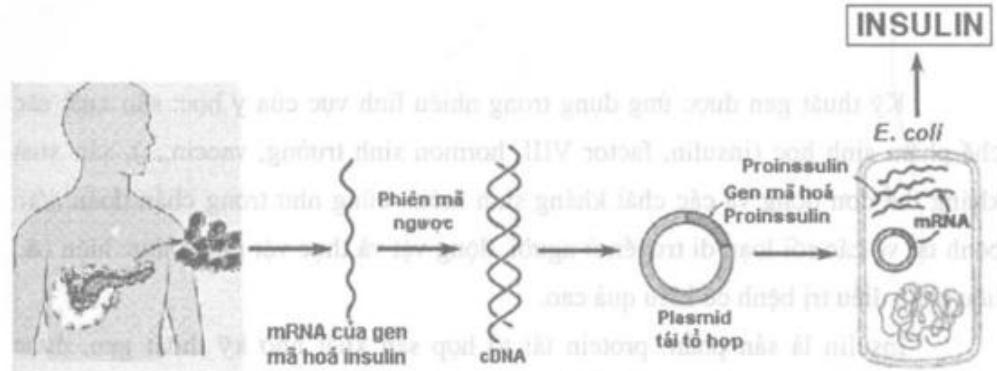
#### Tài liệu tham khảo chính

1. Atienzar F. A., Evenden A. J., Jha A. N., Depledge M. H. (2002), "Use of the random amplified polymorphic DNA (RAPD) assay for the detection of DNA damage and mutations: possible implications of confounding factors", *Biomarkers*, 7, pp. 94-101.
2. Baker S. B., Rugh C. L., Kamalay J. C. (1990), "RNA and DNA isolation from recalcitrant plant tissue", *Biotechniques*, 9, pp. 268-272.
3. Chen X. H., Niu Y. C., Hu B. Z. (2004), "Identification of RAPD markers linked to the resistance gene Yr 5 against wheat stripe rust with denaturing PAGE-silver staining", *Yi Chuan Xue Bao*, 31(3), pp. 270-274.
4. Costa JM, Ernault P, Vidaud D, Vidaud M, Meyer D, Lavergne JM. (2000). Fast and efficient mutation detection method using multiplex PCR and cycle sequencing application to hemophilia A. *Hromb Haemost*. 83, 244, 7
5. Fraga J., Rojas L., Sariego I., Sarria C. A. (2002), "Optimization of random amplified polymorphic DNA techniques for its use in genetic studies of *Trichomonas vaginalis* isolates", *Genet. Evolut*, 2, pp. 73-75.

6. Ida A. A., Julie A. P., Rachel A. L., Guijun Y. (2004), "Fingerprinting of cauliflower cultivars using RAPD markers", *Australian Journal of Agricultural Research*, 55(2), pp. 117-124.
7. Khuất Hữu Thanh, (2003). Cơ sở di truyền phân tử và kỹ thuật gen. NXB Khoa học và Kỹ thuật, Hà Nội.
8. Ma R., Yli-Mattila T., Pulli S. (2004), "Phylogenetic relationships among genotypes of worldwide collection of spring and winter ryes (*Secale cereale* L.) determined by RAPD-PCR markers", *Hereditas*, 140(3), pp. 210-221.
9. Nguyễn Đức Thành, Phạm Huy Toàn, Nguyễn Hoàng Anh, Henry. T Nguyen (2000), "Ứng dụng chỉ thị phân tử RAPD và STS trong nghiên cứu đa dạng di truyền và chọn giống ở lúa", *Những vấn đề nghiên cứu cơ bản trong sinh học*, tr. 149-152.
10. Sambrook J. and Russel D.W. (2001). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
11. Watson J. D. và cộng sự (2004), Molecular Biology of the Gene Fifth edition, Pearson Education, Inc., pub. Benjamin Cummings.
12. Weaver, R. F. (2002), *Molecular Biology*, Mc Graw Hill.
13. Wu S. B., Collins G., Sedgley M. (2004), "A molecular linkage map of olive (*Olea europaea* L) based on RAPD, microsatellite, and SCAR markers", *Genome*, 47(1), pp. 26-35.

### PHẦN III

## MỘT SỐ ỨNG DỤNG CHỦ YẾU CỦA KỸ THUẬT GEN



### **Chương 6:**

### **Ứng dụng kỹ thuật gen trong y học và chẩn đoán**

### **Chương 7:**

### **Kỹ thuật gen và các ứng dụng trong thực tiễn sản xuất**

## CHƯƠNG 6

### MỘT SỐ ỨNG DỤNG CHỦ YẾU CỦA KỸ THUẬT GEN

Kỹ thuật gen được ứng dụng trong nhiều lĩnh vực của y học: sản xuất các chế phẩm sinh học (insulin, factor VIII, hormon sinh trưởng, vaccin,...), sản xuất kháng thể đơn dòng và các chất kháng sinh mới... cũng như trong chẩn đoán sớm bệnh tật và các rối loạn di truyền ở người, động vật và thực vật nhằm thực hiện các liệu pháp điều trị bệnh có hiệu quả cao.

Insulin là sản phẩm protein tái tổ hợp sản xuất nhờ kỹ thuật gen, được thương mại hóa sớm nhất (1992). Đến nay, có trên 400 chế phẩm protein tái tổ hợp được sản xuất bằng kỹ thuật gen, trong đó có hơn 50 chế phẩm được phép sử dụng ở Mỹ và Châu Âu. Các chế phẩm protein tái tổ hợp như interferon, interleukin, hormon, enzym, vaccin tái tổ hợp, vaccin DNA... ngày càng được ứng dụng rộng rãi nhiều trong điều trị nhiều loại bệnh cho con người vật nuôi và cây trồng.

Một số chế phẩm protein tái tổ hợp điển hình, được sử dụng trong điều trị nhiều loại bệnh hiểm nghèo cho con người gồm:

- + Phòng chống bệnh do nhiễm virus, tăng cường miễn dịch  
Actimmune (gamma interferon), Genetech Inc.  
Alferon N, Interferon Sciences Inc.
- Betaseron (recombinant interferon beta 1-B)
- + Làm tan các máu cục gây tắc và nghẽn mạch  
Activase (recombinant alteplase), Genentech Inc.
- + Sử dụng trong chữa bệnh do thiếu hụt miễn dịch trẻ em - immunodeficiency).

- + Chữa bệnh loạn đường cơ hay chữa bệnh Gaucher typ 1  
Ceredase (alglucerase), Genzyme Corp.
- Cerezyme (imiglucerase), Genzyme Corp.
- + Chữa bệnh viêm gan B  
Engerix-B, SmithKline Beecham
- Recombivax HB, Merck & Co.
- Intron A (alpha-interferon), Schering-Plough Corp.
- + Phòng chống HIV- AIDS  
EPOGEN (epoetin alfa), Amgen Ltd.
- + Chữa bệnh lùn do thiếu hụt hormon sinh trưởng ở trẻ em  
Humatrop (somatropin), Eli Lilly & Co.
- Protropin (somatrem), Genentech Inc.
- + Chữa bệnh tiểu đường typ I  
Humulin (human insulin), Eli Lilly & Co.
- + Chữa ung thư  
Enzone, Rhone - Poulenc Rorer
- Proleukin IL – 2, Chiron Corp.
- Roferon-A (recombinant alfa-interferon), Hoffman - La Roche

## I. Kỹ thuật gen trong sản xuất các chế phẩm sinh học

### *1. Sản xuất insulin tái tổ hợp*

#### *1. 1. Lược sử nghiên cứu insulin*

Insulin là hormon được sản xuất bởi các tế bào beta ( $\beta$ ) tuyến tụy, có vai trò quan trọng trong chuyển hóa đường gluco trong máu người và động vật. Thiếu hụt insulin gây bệnh tiểu đường, các bệnh về mắt (bệnh glucom, khô giác mạc, mù loà), bệnh về dạ dày (30 - 50%), bệnh tim, gây lở loét do rối loạn chuyển hóa....

Bệnh tiểu đường (Diabet) do thiếu hụt insulin được phát hiện lần đầu tiên năm 1552 trước Công nguyên, ở một bệnh nhân người Hy Lạp. Năm 1921, Frederick Banting và Charles Best phát hiện insulin được sản xuất bởi các tế bào tuyến tụy.

Từ 1922 con người đã biết tách chiết insulin từ tụy bò, tuy lợn để chữa bệnh tiểu đường cho con người. Do insulin người khác insulin của lợn 1 acid amin và khác với insulin của bò 3 acid amin. Sử dụng insulin từ tụy bò và tụy lợn để chữa bệnh cho người bị tiểu đường có hiệu quả không cao, sau một thời gian điều trị nhất định hiệu quả điều trị giảm dần, có thể do phản ứng loại thải protein lạ ở người bệnh.

Cấu trúc phân tử insulin người được phát hiện năm 1926. Năm 1955, Frederick Sanger và cộng sự bằng kỹ thuật giải trình tự đã xác định trật tự sắp xếp của 51 acid amin trong phân tử insulin người (chuỗi peptid đầu tiên được giải trình tự, Frederick Sanger được nhận giải Nobel năm 1962 về xây dựng cấu trúc protein).

Năm 1963 - 1965 insulin được tổng hợp nhân tạo bằng con đường hoá học, tuy nhiên giá thành quá đắt không phù hợp trong chữa bệnh.

Năm 1967 phát hiện cấu trúc phân tử preinsulin (Steiner và Chance, 1967).

Năm 1969, Crowfoot - Hodgkin phát hiện kích thước 3 dạng cấu trúc của insulin.

**Bảng 6. 1. Sự khác nhau của insulin ở người so với insulin ở bò và lợn**

Insulin	Chuỗi A		Chuỗi B
	a. a thứ 8	a. a thứ 10	a..a thứ 30
Người	Thr	Ile	Thr
Bò	Ala	Val	Ala
Lợn	Thr	Ile	Ala

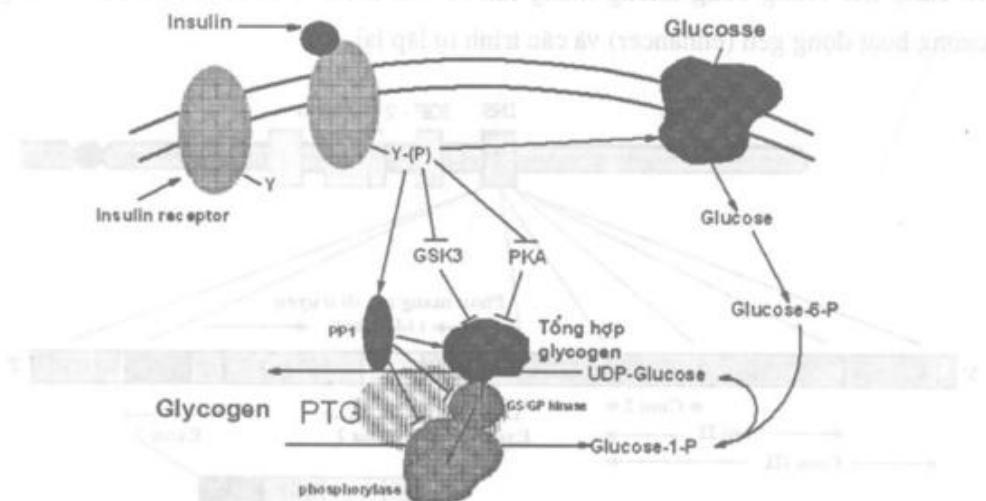
Năm 1978 Herbert Boyer (Công ty Genetech) thành công sản xuất insulin tái tổ hợp trong vi khuẩn *E. coli*.

Năm 1982, sản phẩm insulin người do hãng Novo sản xuất được phép lưu hành ở Châu Âu. Năm 1983, sản phẩm insulin người được thương mại hóa ở Mỹ.

Năm 1986 sản phẩm thương mại đầu tiên gọi là Humulin do Công ty Eli Lily và Genetech sản xuất, được sử dụng ứng dụng trong chữa bệnh cho con người. Năm 1987 thành công tổng hợp nhân tạo insulin trong tế bào nấm men.

Đến nay có rất nhiều chế phẩm insulin với nhiều tên thương mại khác nhau được bán trên thị trường, góp phần điều trị bệnh, bảo vệ sức khoẻ con người.

Insulin có vai trò quan trọng, giúp điều chỉnh nồng độ đường trong máu ở mức thích hợp. Khi hàm lượng glucose trong máu cao, insulin đóng vai trò các chìa khoá cho phép glucose từ máu qua màng tế bào chất vào trong tế bào, đồng thời insulin xúc tác quá trình chuyển hoá glucose trong gan thành các chất dự trữ glycogen và các chất béo. Ngược lại, khi hàm lượng glucose trong máu thấp, insulin giữ vai trò chuyển hoá glycogen dự trữ trong gan thành glucose đưa vào máu, không cho glucose từ máu chuyển vào tế bào.



Hình 6.1. Vai trò của insulin trong sự chuyển hoá đường glucose

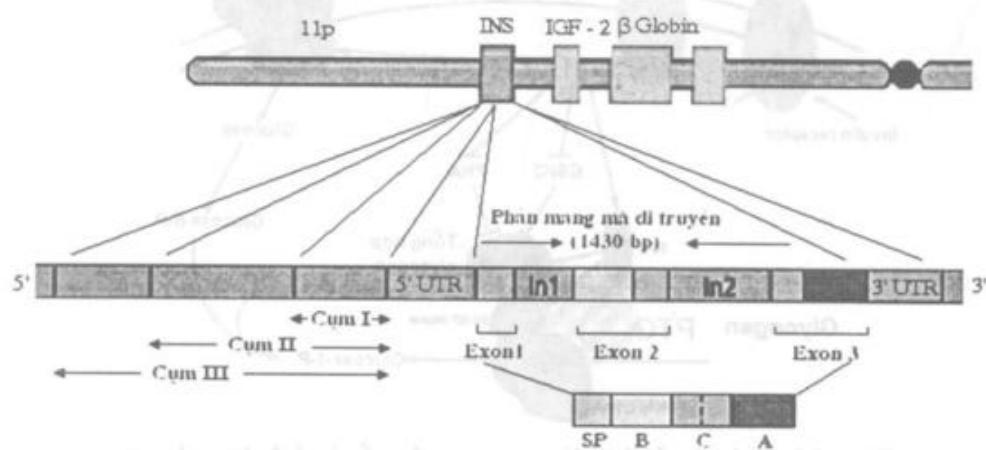
Thiếu hụt insulin trong cơ thể gây nên bệnh tiểu đường và nhiều bệnh về mắt, bệnh dạ dày...

Bệnh tiểu đường gồm 2 dạng chính gọi là typ I và typ II. Bệnh tiểu đường typ I thường gặp ở trẻ em, do trong cơ thể gen insulin bị hỏng, dẫn đến hàm lượng insulin quá ít hoặc thiếu hụt hoàn toàn. Bệnh tiểu đường typ II thường gặp ở người lớn do bị bệnh béo phì hoặc thiếu một số loại enzym chuyển hóa trong cơ thể. Trường hợp tiểu đường typ II, gen insulin hoạt động yếu tạo ra trong cơ thể hàm lượng insulin dưới mức bình thường, hoặc quá ít. Hiện nay có khoảng trên 2 tỉ người trưởng thành bị tiểu đường (10% typ I và 90% typ II).

### 1.2. Cấu trúc gen mã hoá insulin ở người

Nghiên cứu bộ gen người cho thấy gen mã hoá insulin (INS) nằm ở nhiễm sắc thể số 11, trên cánh ngắn (11p) gần gen IGF và gen  $\beta$ -globin. Gen mã hóa insulin người gồm 2 phần: vùng mang mã di truyền (coding region) và vùng không mang mã (no coding region).

Vùng không mang mã di truyền chia làm 3 cụm gen kí hiệu class I, class II và class III. Trong vùng không mang mã có các trình tự điều hoà, trình tự tăng cường hoạt động gen (enhancer) và các trình tự lặp lại.



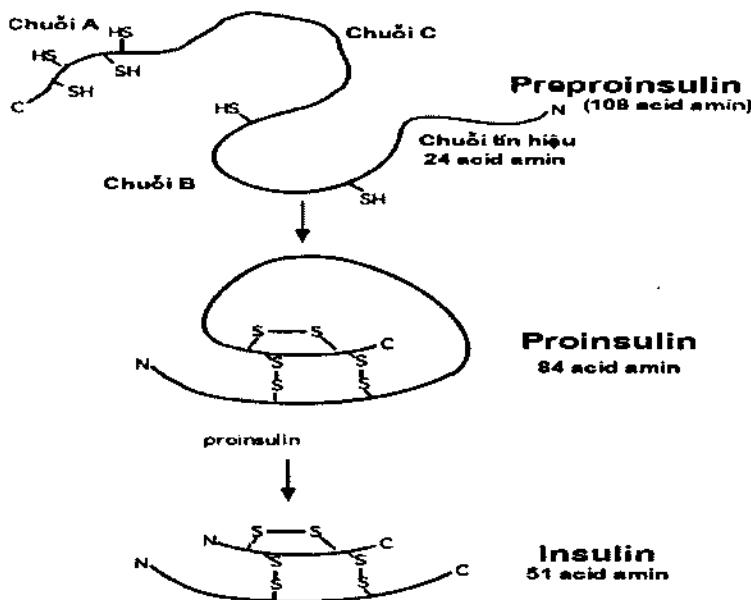
Hình 6.2. Cấu trúc gen mã hoá insulin ở người

Vùng mang mã di truyền có kích thước 1430 bp. Trong vùng mang mã di truyền có 2 intron, intron thứ nhất (In1) nằm giữa đoạn trình tự mã hoá chuỗi peptid

tín hiệu và chuỗi peptid B, intron thứ 2 (In2) nằm giữa đoạn gen mang mã di truyền mã hoá chuỗi peptid C.

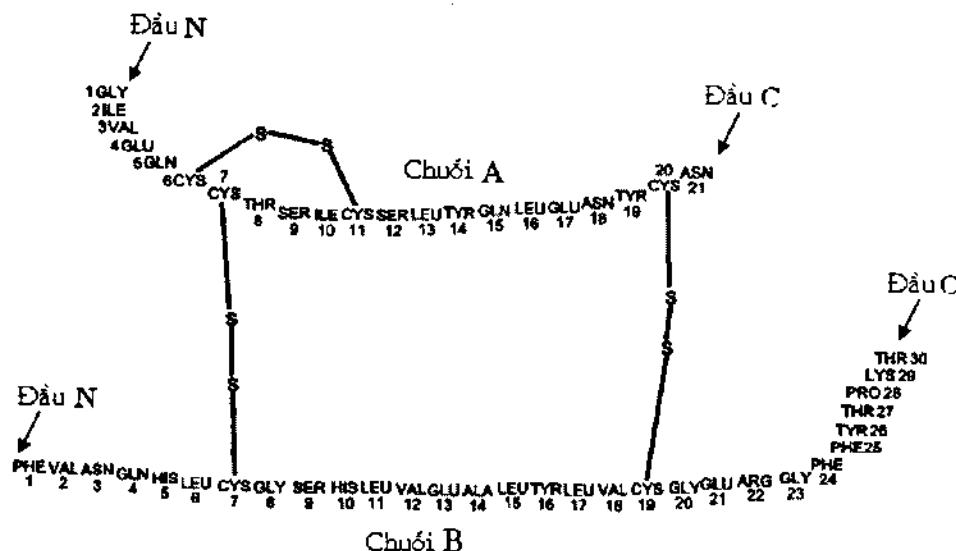
Đoạn gen mang mã di truyền mã hoá 4 chuỗi peptid trong cấu trúc của phân tử insulin: chuỗi peptid tín hiệu (SP- signal peptide), chuỗi peptid B, chuỗi peptid C và chuỗi peptid A.

Trong cơ thể người insulin được tổng hợp qua một chuỗi các biến đổi, đầu tiên phân tử preproinsulin được tổng hợp. Preproinsulin là chuỗi peptid dài gồm 108 acid amin, gồm 4 chuỗi peptid ngắn: chuỗi peptid tín hiệu (SP), chuỗi peptid B, chuỗi peptid C và chuỗi peptid A. Sau đó, preproinsulin được đưa tới mạng lưới nội chất, ở đây bị cắt mất chuỗi tín hiệu, cấu trúc biến đổi tạo nên phân tử proinsulin có 84 acid amin.



Hình 6. 3. Sơ đồ biến đổi cấu trúc preproinsulin

Trong mạng lưới nội chất, phân tử proinsulin được cắt bỏ chuỗi C và hình thành các liên kết disulfit, chuyển thành dạng insulin có hoạt tính gồm 51 acid amin (chuỗi A có 21 acid amin và chuỗi B có 30 acid amin).



Hình 6. 4. Trình tự acid amin trong phân tử insulin

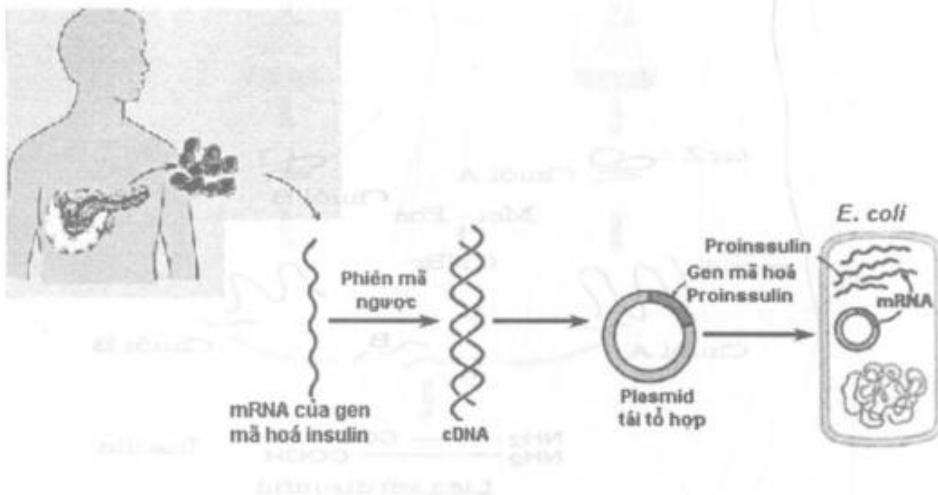
### 1. 3. Công nghệ sản xuất insulin tái tổ hợp

Hiện nay công nghệ sản xuất insulin tái tổ hợp đã hoàn thiện. Insulin được sản xuất ở qui mô công nghiệp với nhiều các loại tế bào chủ khác nhau (vì khuẩn, nấm men ...). Nhiều loại sản phẩm insulin tái tổ hợp, với các tên thương mại khác nhau Humulin®; Humalog® (Eli Lilly); Novolin® (Novo Nordisk)..., ngày càng được ứng dụng rộng rãi trong điều trị bệnh tiểu đường cho con người.

Nguyên lý chung của phương pháp sản xuất insulin tái tổ hợp gồm các bước cơ bản sau:

- Tách chiết và tinh sạch RNA thông tin mã hóa preproinsulin từ các tế bào beta tuyến tụy của người.
- Thực hiện kỹ thuật phiên mã ngược tạo phân tử cDNA mạch kép.
- Sử dụng enzym giới hạn thích hợp cắt bỏ đoạn gen mã hóa chuỗi peptid C, đồng thời tách dòng riêng từng đoạn gen mã hóa chuỗi peptid A (đoạn gen A) và đoạn gen mã hóa chuỗi peptid B (đoạn genB).

- Tạo các vectơ biểu hiện gen tái tổ hợp mang đoạn gen A và vectơ tái tổ hợp mang đoạn gen B. Trong nhiều trường hợp để biểu hiện gen mã hoá insulin trong vi khuẩn hoặc nấm men, người ta thường sử dụng vectơ biểu hiện có gen chỉ thị *lac Z*.



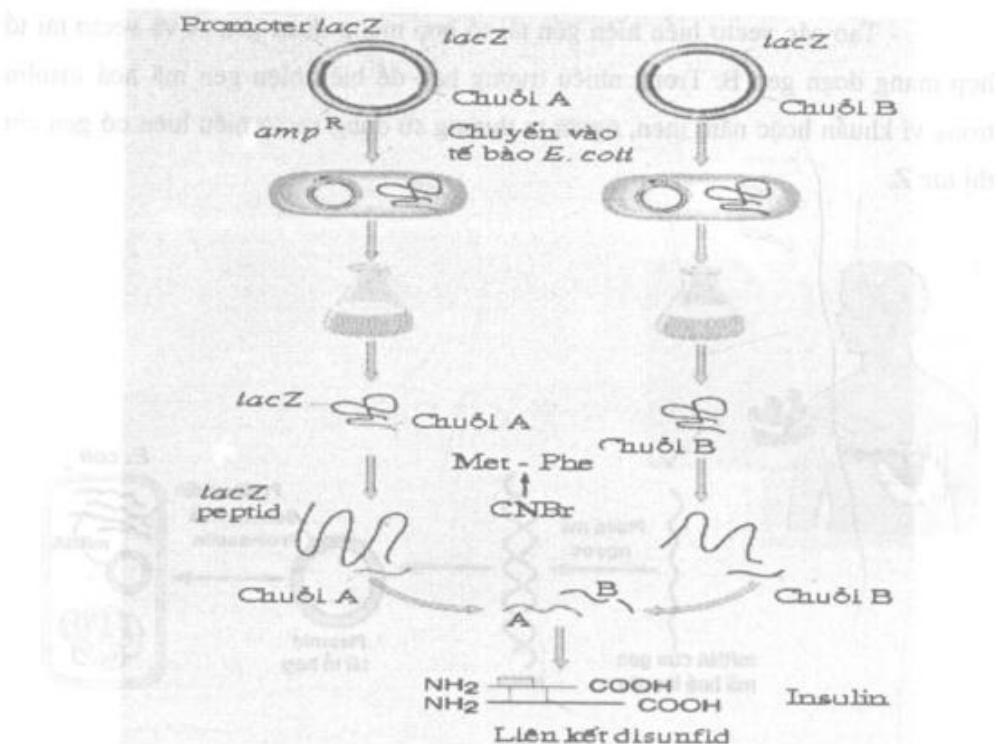
Hình 6.5. Sơ đồ tạo plasmid tái tổ hợp và biến nạp vào *E. coli*

- Biến nạp riêng mỗi loại vectơ tái tổ hợp mang đoạn gen A (hoặc đoạn gen B) vào các dòng tế bào *E. coli* khác nhau.

- Thực hiện quá trình lén men thu sinh khói, tạo các sản phẩm protein lai. Trong một dòng tế bào *E. coli* có sự tạo thành các phân tử protein lai: chuỗi A - *lac Z*, một dòng tạo nên protein lai: chuỗi B - *lac Z*.

- Thu riêng từng loại protein lai và thực hiện các phản ứng với cyanua bromid (CNBr), cắt liên kết giữa các acid amin cuối cùng của chuỗi peptid với acid amin đầu tiên của enzym  $\beta$  - Galactosidase (Phe - Met).

- Tách và tinh sạch riêng biệt các chuỗi peptid A và peptid B, sau đó tạo hỗn hợp các chuỗi peptid A với chuỗi peptid B trong điều kiện thích hợp, hình thành các liên kết disulfit giữa các acid amin cystein để thu insulin thành phẩm.



Hình 6.6. Sơ đồ công nghệ sản xuất insulin tái tổ hợp trong *E. coli*

Hiện nay, một số hãng chuyên sản xuất protein tái tổ hợp trên thế giới (Avigenic, TranXenogen, PPL của Viện Roslin...) đã chuyển được gen mã hoá insulin người vào số loài động vật như dê, cừu, thỏ, chuột và gia cầm tạo ra các động vật cho các sản phẩm (sữa, trứng) có hàm lượng insulin cao.

## 2. Sản xuất hormon sinh trưởng người tái tổ hợp

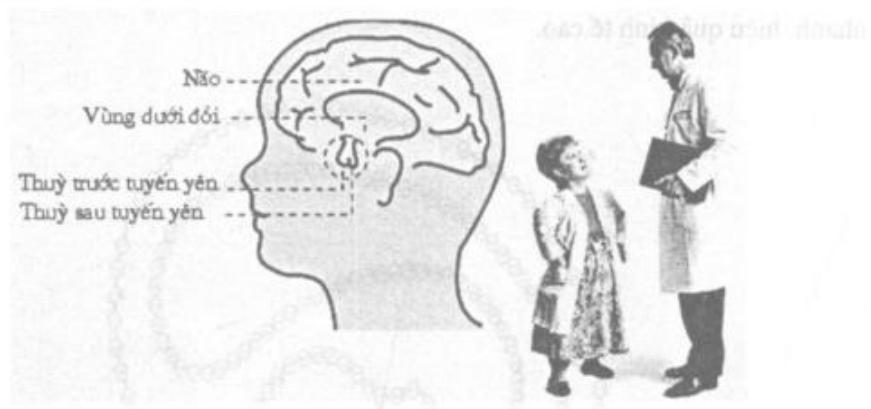
Hormon sinh trưởng người (hGH - Growth Hormone) là loại hormon được sản xuất ở các tế bào tuyến yên trong não, có vai trò điều hoà quá trình sinh trưởng và phát triển của xương, cơ bắp... thiếu hụt hGH làm cho cơ thể chậm phát triển, hoặc có thể gây hội chứng lùn侏儒症 ở trẻ em.

## 2. 1. Lược sử nghiên cứu Hormon sinh trưởng người (hGH)

Hormon sinh trưởng người được phát hiện năm 1920. Năm 1950, đã tách chiết được hGH từ tuyến yên của não người chết với dung môi acetone.

Từ năm 1956, hormon sinh trưởng tách được chiết từ tuyến yên não bò, não cừu... để chữa bệnh lùn cho một số người bệnh. Hormon sinh trưởng cừu được sử dụng điều trị bệnh lùn ở trẻ em (để có được 5 mg hormon sinh trưởng cừu cần tách chiết từ 500.000 não cừu), giá thành chữa bệnh rất đắt.

Năm 1972, lần đầu tiên cấu trúc phân tử của hGH được xác định đầy đủ. Năm 1985, Công ty Genetech sản xuất thành công hormon sinh trưởng người (hGH) tái tổ hợp trong vi khuẩn *E. coli* bằng kỹ thuật gen với tên thương mại là Protropin.



*Hình 6. 7. Tuyến yên sản xuất hGH và bệnh lùn do thiếu hGH*

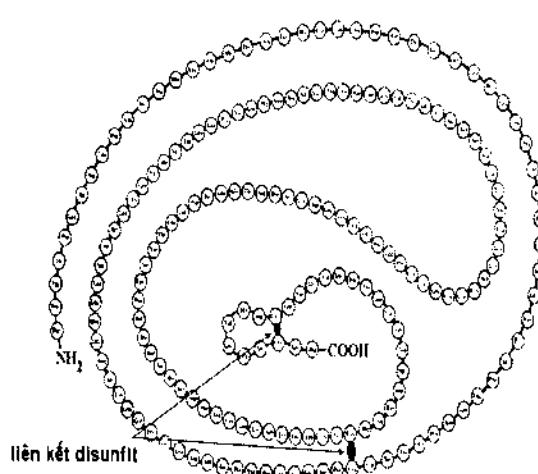
Hormon sinh trưởng người tái tổ hợp hiện nay chủ yếu được sản xuất bằng phương pháp tái tổ hợp gen. Có nhiều loại sản phẩm thương mại khác nhau được ứng dụng trong điều trị bệnh cho con người hoặc giúp cho sự phát triển cơ bắp của các vận động viên thể hình: Genotropin, Humantrope, Norditropin, Nutropin, ProLease, Saizen, Serostim ....

Ở nhiều nước trên thế giới hormon sinh trưởng người tái tổ hợp được sử

được 5 mg hGH chỉ cần thực hiện lên men 15 giờ các chủng vi khuẩn tái tổ hợp trong bình lên men 5 lít, tạo nên sản phẩm hormon sinh trưởng người có giá thành rẻ hơn rất nhiều so với tác chiết từ não động vật, đồng thời tránh được sự lây truyền các bệnh truyền nhiễm, bệnh bò điên...

Hormon sinh trưởng người là protein có kích thước nhỏ, gồm 191 acid amin với khối lượng phân tử 22.125 kD, trong cấu trúc phân tử có sự hình thành 2 liên kết disulfide giữa các acid amin cystein.

Hormon sinh trưởng người có vai trò quan trọng trong sự phát triển bình thường của người, thiếu hGH trẻ em bị hội chứng lùn. Hiện nay, ngoài việc chuyển gen mã hoá hormon sinh trưởng người vào vi sinh vật để sản xuất hGH tái tổ hợp, gen mã hoá hormon sinh trưởng người, bò... đã được chuyển vào nhiều loài động vật (cá hồi, lợn, cừu...) tạo các giống động vật chuyển gen có khả năng sinh trưởng nhanh, hiệu quả kinh tế cao.

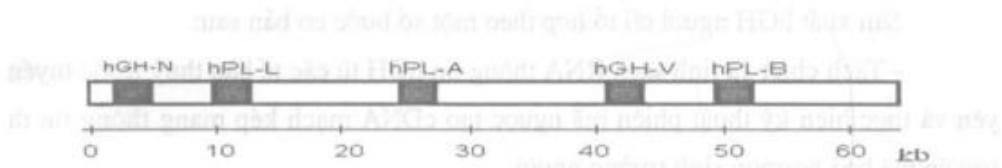


**Hình 6. 8. Cấu trúc phân tử hormon sinh trưởng người (hGH)**

## 2. 2. Cấu trúc gen mã hoá hormon sinh trưởng người

Hormon sinh trưởng người do một họ gen có 5 gen mã hoá, các gen nằm trên cánh dài của nhiễm sắc thể số 17 của người (17q23.24) kích thước khoảng 66

kb. Gen hGH - N kích thước 22 kb mã hoá hGH, gen hGH - V có kích thước khoảng 20 kb. Các gen hPL - L, hPL - A và hPL - B mã hoá prolactin.



Hình 6.9. Cấu trúc cụm gen mã hoá hormon sinh trưởng người

Gen mã hoá hGH bị đột biến hoặc rối loạn, dẫn đến tình trạng thiếu hụt hormon sinh trưởng người, gây nên hội chứng lùn bẩm sinh di truyền qua các thế hệ trong gia đình.



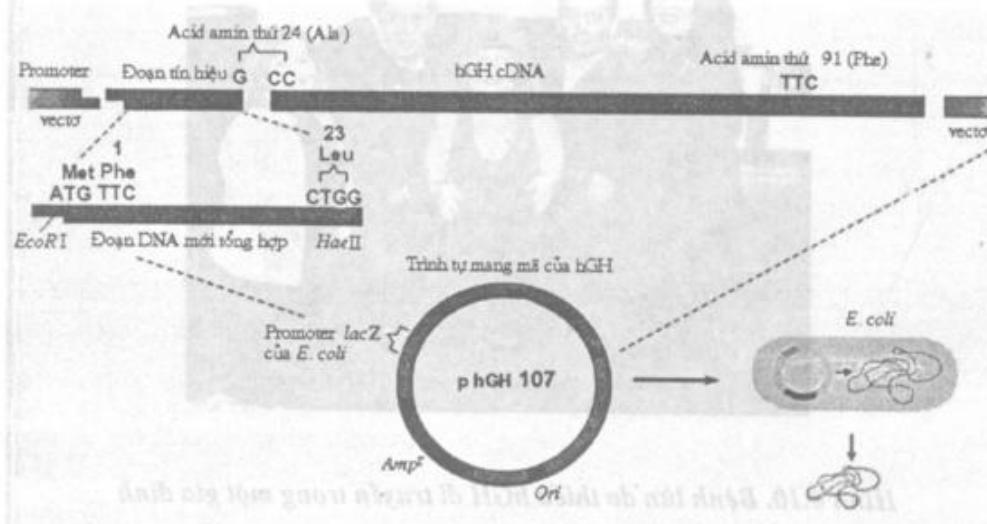
Hình 6.10. Bệnh lùn do thiếu hGH di truyền trong một gia đình

Gen hGH có đoạn trình tự mã hoá chuỗi peptid tín hiệu không tương thích promoter của *E. coli*, do đó biểu hiện gen hGH tái tổ hợp trong tế bào *E. coli* rất yếu hoặc không hoạt động. Để có thể tạo hGH tái tổ hợp trong tế bào *E. coli*, cần thay đổi cấu trúc đoạn gen mã hoá peptid tín hiệu trước khi tách dòng gen hGH.

### 2.3. Công nghệ sản xuất hormon sinh trưởng người tái tổ hợp

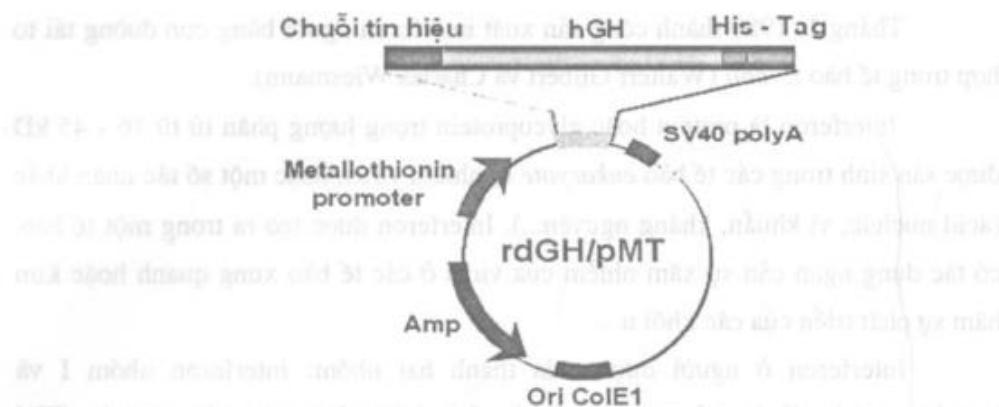
Sản xuất hGH người tái tổ hợp theo một số bước cơ bản sau:

- Tách chiết và tinh sạch RNA thông tin hGH từ các tế bào thuỷ trước tuyến yên và thực hiện kỹ thuật phiên mã ngược tạo cDNA mạch kép mang thông tin di truyền mã hoá hormon sinh trưởng người.
- Sử dụng các enzym giới hạn cắt bỏ đoạn gen mã hoá 23 - 24 acid amin của chuỗi peptid tín hiệu. Người ta sử dụng enzym giới hạn *EcoR* I cắt bỏ đoạn trình tự tín hiệu của gen hGH, sau đó tổng hợp nhân tạo một đoạn DNA mã hoá 23 - 24 acid amin đầu tiên (thay đổi trình tự gen nhưng không thay đổi trình tự acid amin). Đoạn DNA tái tổ hợp mã hoá chuỗi peptid tín hiệu được nối với đoạn còn lại của gen hGH để thực hiện quá trình tái tổ hợp lại tạo gen mã hoá hGH.
- Chuyển gen mã hoá hGH tái tổ hợp vào các vectơ biểu hiện thích hợp.



Hình 6.11. Công nghệ sản xuất hGH tái tổ hợp trong tế bào *E. coli*

- Biến nạp vectơ tái tổ hợp vào *E. coli*, nuôi cấy trong những điều kiện thích hợp thu sinh khối tế bào vi khuẩn *E. coli*, tách chiết và tinh sạch hGH.



**Hình 6.12. Vector biểu hiện rdGH trong sản xuất hGH**

Tùy theo hệ thống tế bào biểu hiện trong công nghệ sản xuất hormon sinh trưởng người tái tổ hợp (vi khuẩn, baculovirus, nấm men, tế bào trứng chuột đất vàng Trung Quốc - CHO...), có thể sử dụng các loại vectơ biểu hiện khác nhau. Trong sản xuất hGH người tái tổ hợp thường sử dụng các vectơ biểu hiện là plasmid (phGH 107) để biểu hiện trong tế bào *E. coli*. Biểu hiện gen hGH trong tế bào sinh vật bậc cao thường sử dụng vectơ biểu hiện rdGH/ pMT ...

### 3. Sản xuất interferon tái tổ hợp

#### 3. 1. Lược sử nghiên cứu interferon (IFN)

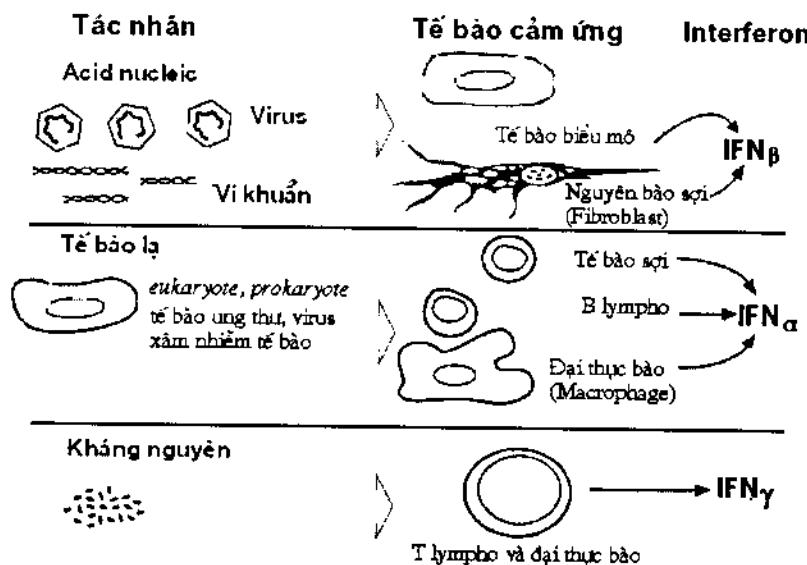
Từ năm 1930, người ta đã phát hiện hiện tượng khi gia súc bị nhiễm bệnh do một loại virus gây nên, ít bị nhiễm bệnh do các loại virus khác. Năm 1957, Alick Isaacs (Anh) và Jean Lindenmann (Thụy Sĩ) khi nuôi cấy các tế bào phổi gà phát hiện sự hình thành một loại protein mới gây ức chế tái bản của virus, gọi là interferon (IFN).

Trước 1980, sản xuất IFN từ tế bào bạch cầu người, bằng cách gây nhiễm virus tế bào bạch cầu sau đó nuôi trong phôi gà để các tế bào bạch cầu sinh interferon nên giá thành interferon rất đắt.

Tháng 1. 1980 thành công sản xuất interferon người bằng con đường tái tổ hợp trong tế bào *E. coli* (Walter Gilbert và Charles Wiesmann).

Interferon là protein hoặc glycoprotein trọng lượng phân tử từ 16 - 45 kD được sản sinh trong các tế bào *eukaryote* bị nhiễm virus, hoặc một số tác nhân khác (acid nucleic, vi khuẩn, kháng nguyên...). Interferon được tạo ra trong một tế bào, có tác dụng ngăn cản sự xâm nhiễm của virus ở các tế bào xung quanh hoặc kiềm hãm sự phát triển của các khối u ...

Interferon ở người được chia thành hai nhóm: interferon nhóm I và interferon nhóm II. Interferon nhóm I gồm 3 loại interferon: interferon alpha (IFN  $\alpha$ ), interferon beta (IFN  $\beta$ ) và interferon omega (IFN  $\omega$ ). Các interferon thuộc nhóm I tương đối bền trong điều kiện acid. Interferon nhóm II gồm một loại interferon gamma (IFN  $\gamma$ ) kém bền trong môi trường acid.



Hình 6. 13. Các tế bào sản sinh loại interferon chủ yếu

Interferon alpha (IFN  $\alpha$ ) là loại protein có 166 acid amin với khoảng 18 loại acid amin khác nhau (IFN  $\alpha$  2 có 165 acid amin). Trọng lượng phân tử của

interferon alpha khoảng 19, 4 kD, IFN  $\alpha$  được sản sinh trong các tế bào bạch cầu người.

Interferon beta (IFN  $\beta$ ) là protein có 187 acid amin, trọng lượng phân tử khoảng 20, 0 kD. Interferon beta được sản sinh từ các tế bào nguyên bào sợi (fibroblast) và tế bào biểu bì.

Interferon omega (IFN  $\omega$ ) là protein gồm 195 acid amin. Interferon gamma (IFN  $\gamma$ ) hay interferon miễn dịch là protein gồm 220 acid amin, trọng lượng phân tử 17 kD, được tạo ra từ các tế bào đại thực bào lympho T và lympho T <sub>KL</sub> (lympho T killer).

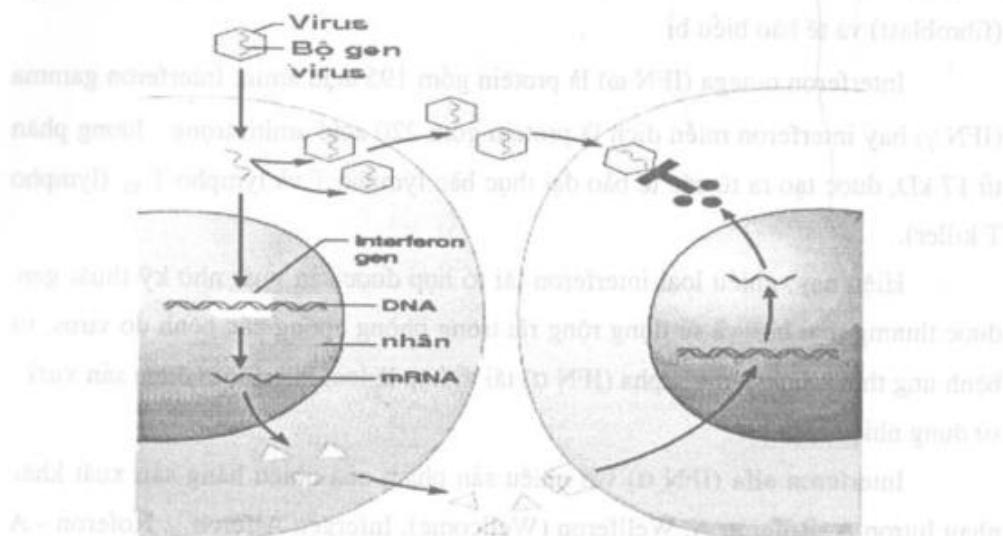
Hiện nay, nhiều loại interferon tái tổ hợp được sản xuất nhờ kỹ thuật gen, được thương mại hoá và sử dụng rộng rãi trong phòng chống các bệnh do virus, và bệnh ung thư... Interferon alpha (IFN  $\alpha$ ) tái tổ hợp là loại interferon được sản xuất sử dụng nhiều nhất.

Interferon alfa (IFN  $\alpha$ ) với nhiều sản phẩm của nhiều hãng sản xuất khác nhau Intron A, Roferon A, Wellferon (Wellcome), Infergen Alferon... Roferon - A (Interferon alfa - 2a, Roche sản xuất) là sản phẩm được tái tổ hợp trong *E. coli*. Interferon alfa - 2a gồm 165 acid amin trong đó có 23 acid amin lizin (lysine), trọng lượng phân tử khoảng 19. 000 Daltons (~19 kDa). Interferon alfa - 2b (IFN  $\alpha$  2b) với nhiều sản phẩm khác nhau: Intron A (Schering Plough), Laferon (Nga), Inrec (Cuba), Superferon (Việt Nam)... Interferon alfa được sử dụng trong phòng chống các bệnh AIDS, ung thư sarcoma, bệnh nhiễm virus viêm gan B, viêm gan C, virus herpes và HIV....

Interferon beta (IFN  $\beta$ ) tái tổ hợp gồm nhiều tên thương mại khác nhau: IFN -  $\beta$  1a với tên gọi Avonex<sup>TM</sup>, gồm 166 acid amin được tái tổ hợp trong tế bào trứng chuột đất vàng Trung Quốc (CHO). IFN -  $\beta$  1b với tên gọi Betaseron gồm 165 acid amin được tái tổ hợp trong tế bào *Escherichia coli*, Refib gồm 166 acid amin.... Interferon beta (IFN  $\beta$ ) được sử dụng trong điều trị các bệnh ung thư phổi, ung thư tế bào sắc tố, và các bệnh do nhiễm virus.

Interferon gamma (IFN  $\gamma$ ) tái tổ hợp điển hình là interferon-gamma-1b với

tên thương mại Actimmune, được tái tổ hợp trong tế bào *Escherichia coli*. Interferon gamma được sử dụng trong điều trị các bệnh "Chronic granulomatous", bệnh rối loạn chuyển hoá do nhiễm virus, các bệnh do nhiễm vi khuẩn mycobacterium ...



**Hình 6.14. Cơ chế sản sinh interferon**

### 3.2. Gen mã hoá interferon ở người và cơ chế sinh IFN trong tế bào

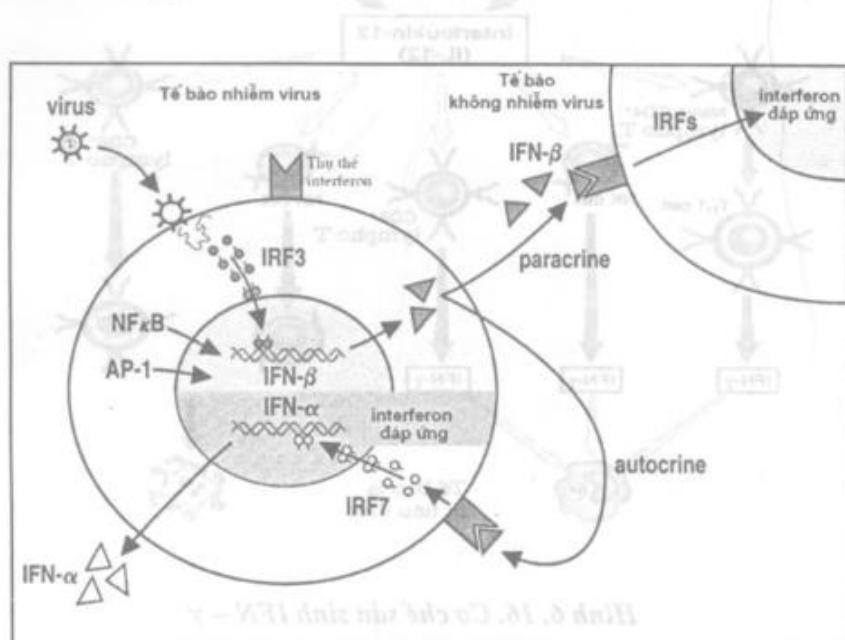
Khi tế bào bị nhiễm virus, sự xâm nhiễm của virus kích hoạt các gen mã hoá interferon hoạt động tạo nên interferon trong tế bào. Interferon mới được tạo thành chuyển qua gian bào sang các tế bào xung quanh, kích thích các tế bào xung quanh sản sinh các kháng thể ngăn cản sự xâm nhiễm của các loại virus khác.

Trong tế bào người, interferon alfa (IFN α) được mã hoá bởi 22 gen nằm trên cánh ngắn của NST số 9. Interferon beta (IFN β) do 1 gen nằm trên NST số 2 mã hoá. Các loại interferon khác nhau, được sản xuất trong các loại tế bào khác nhau theo những cơ chế đặc trưng riêng.

Interferon beta (IFN β) được sản sinh khi có các tác nhân xâm nhiễm tế

bào. Sau khi được tạo thành trong tế bào, IFN β được giải phóng ra ngoài tế bào và

hoạt động theo hai hướng. Hướng thứ nhất: IFN  $\beta$  được chuyển tới các tế bào xung quanh, kích hoạt các tế bào xung quanh sản xuất kháng thể chống lại sự xâm nhiễm của virus. Hướng thứ hai: IFN -  $\beta$  ra khỏi tế bào, tác động tới các thụ thể trên màng tế bào sinh ra nó để kích hoạt tế bào sản sinh interferon alfa (IFN -  $\alpha$ ) chống lại các tác nhân virus xâm nhiễm tế bào.

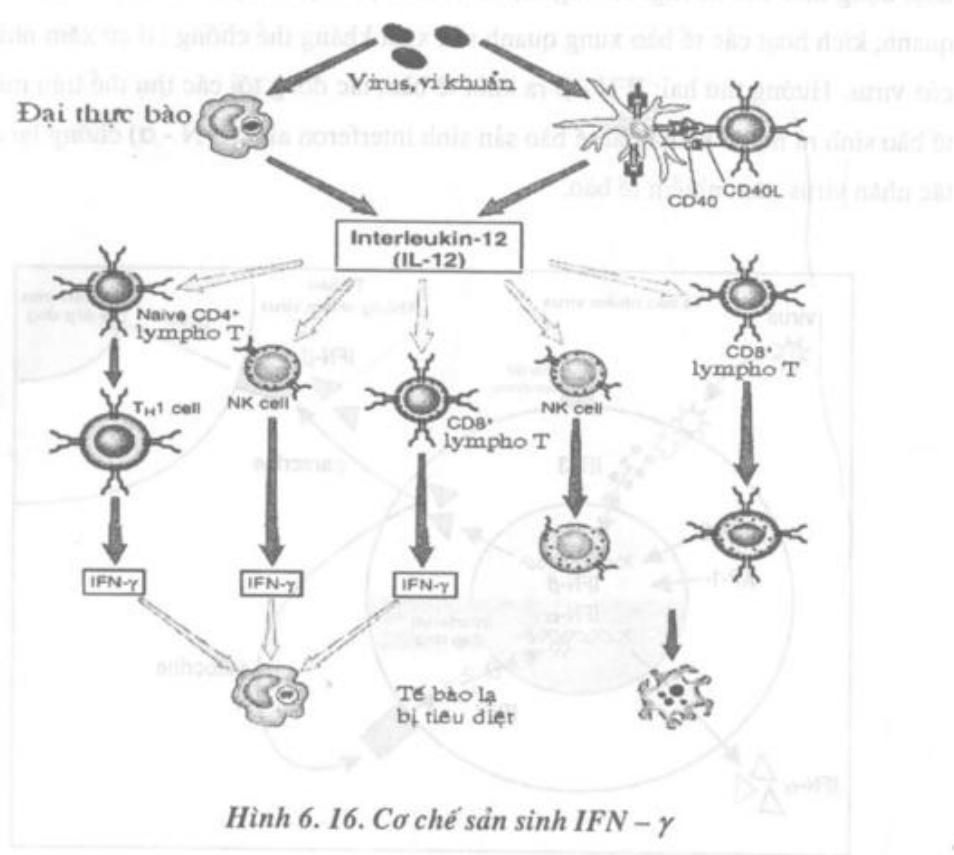


Hình 6.15. Cơ chế sản sinh IFN -  $\beta$  và IFN -  $\alpha$

Interferon gamma (IFN  $\gamma$ ) do 1 gen nằm trên NST số 12 của người mã hoá.

Gen mã hoá IFN  $\gamma$  gồm 4 exon xen kẽ giữa các intron. Interferon gamma được sản sinh trong các tế bào lympho T (tế bào T<sub>H</sub>, T<sub>KL</sub>, T<sub>CDR</sub>...).

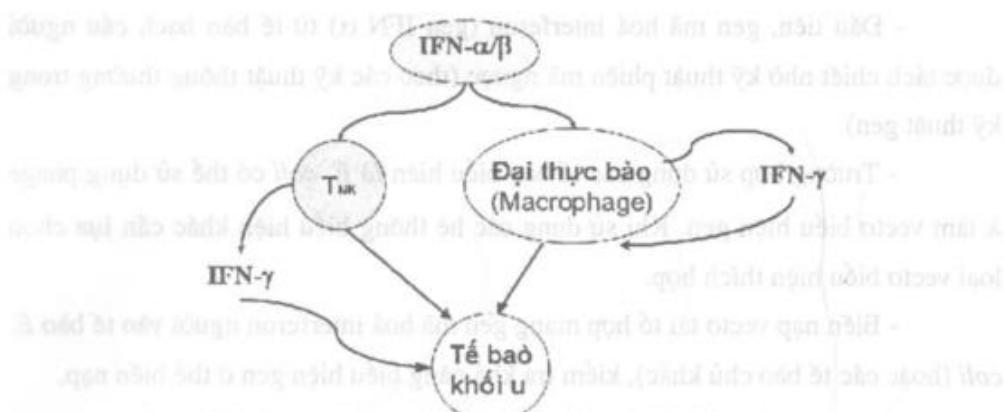
Hiện nay, interferon được ứng dụng rộng rãi trong thực tiễn. IFN có vai trò quan trọng trong miễn dịch tế bào và tác dụng phòng chống nhiễm virus. IFN có tác dụng hoạt hoá các enzym endonuclease gây thoái hoá RNA virus, ức chế phiến mã của virus làm hạn chế quá trình lắp ghép các virion, ức chế sinh trưởng phát triển của virus...



Hình 6. 16. Cơ chế sản sinh IFN -  $\gamma$

### 1. 3. Vai trò của interferon trong điều trị các bệnh do nhiễm virus

Interferon không chỉ được sử dụng phổ biến ở nước ngoài, mà nhiều cơ sở y tế trong nước đã sử dụng interferon có hiệu quả cao trong điều trị viêm gan virus, và ứng dụng trong điều trị ung thư (Bệnh viện Chợ Rẫy TP. Hồ Chí Minh, Bệnh viện Bạch Mai, Bệnh viện K Hà Nội....). Gần đây, superferon do Việt Nam sản xuất được sử dụng trong điều trị bệnh ung thư, phòng chống bệnh do nhiễm virus có hiệu quả cao.



Hình 6.17. Vai trò của IFN α và IFN β đối với tế bào khối u

Tác dụng kiểm soát tế bào ung thư của IFN α và IFN β, do IFN α và IFN β tác động tới các đại thực bào (macrophage) và tế bào lympho T<sub>KL</sub> sản sinh interferon γ tăng khả năng miễn dịch tự nhiên, hạn chế sự phát triển và phân chia của tế bào khối u.

### 3.3. Công nghệ sản xuất interferon tái tổ hợp

Sản xuất interferon tái tổ hợp có thể sử dụng các vectơ biểu hiện khác nhau, biểu hiện trong các loại tế bào chủ là vi khuẩn *E. coli*, baculovirus hoặc tế bào nấm men ....

+ **Sản xuất IFN -α**

IFN-α ở người được sản xuất gây nhiễm nhân tạo virus vào tế bào bạch cầu người. Nhiều loại virus được sử dụng để sản xuất IFN-α (virus Sendai, virus Newcastle, phage λ... các loại virus này kích thích tế bào bạch cầu sản xuất các IFN-α, một phần nhỏ IFN-β và IFN-ω cũng được tạo thành.

- Đầu tiên, gen mã hoá interferon (gen IFN α) từ tế bào bạch cầu người được tách chiết nhờ kỹ thuật phiên mã ngược (theo các kỹ thuật thông thường trong kỹ thuật gen).

- Trường hợp sử dụng các tế bào biểu hiện là *E. coli* có thể sử dụng phage λ làm vectơ biểu hiện gen. Khi sử dụng các hệ thống biểu hiện khác cần lựa chọn loại vectơ biểu hiện thích hợp.

- Biến nạp vectơ tái tổ hợp mang gen mã hoá interferon người vào tế bào *E. coli* (hoặc các tế bào chủ khác), kiểm tra khả năng biểu hiện gen ở thể biến nạp.

- Lên men thu sinh khối tế bào, tách chiết interferon và tinh sạch.

- Thêm các phụ gia (NaCl, KH<sub>2</sub>PO<sub>4</sub>, dextran...), nghiên cứu điều kiện giữ hoạt tính và đóng gói và và thử nghiệm lâm sàng để kiểm tra tính an toàn của sản phẩm ...

#### **+ Sản xuất IFN -β (Betaferon)**

Interferon beta IFN-β được sản xuất chủ yếu ở các nguyên bào sợi của người. Nguyên lí chung của công nghệ sản xuất IFN-β gồm một số bước cơ bản sau:

- Tách chiết và thu nhận gen mã hoá IFN-β từ các tế bào nguyên bào sợi ở người, sau đó cản biến đổi gen (modifying the gene) để tạo sản phẩm mã hoá là IFN-β -1b.

- Tạo vectơ tái tổ hợp bằng cách chèn gen mã hoá cho IFN-β-1b vào vectơ tách dòng thích hợp.

- Biến nạp vectơ tái tổ hợp vào tế bào chủ (thường sử dụng *E. coli*), nuôi cấy trong các thiết bị lên men thu sinh khối, sau đó tách chiết interferon IFN-β-1b và tinh sạch.

- Thêm các phụ gia, các chất ổn định (chất ổn định thường dùng là đường mannitol, albumin...), sau đó tinh chế, tiệt trùng, nghiên cứu điều kiện giữ hoạt tính và đóng gói và và thử nghiệm lâm sàng...

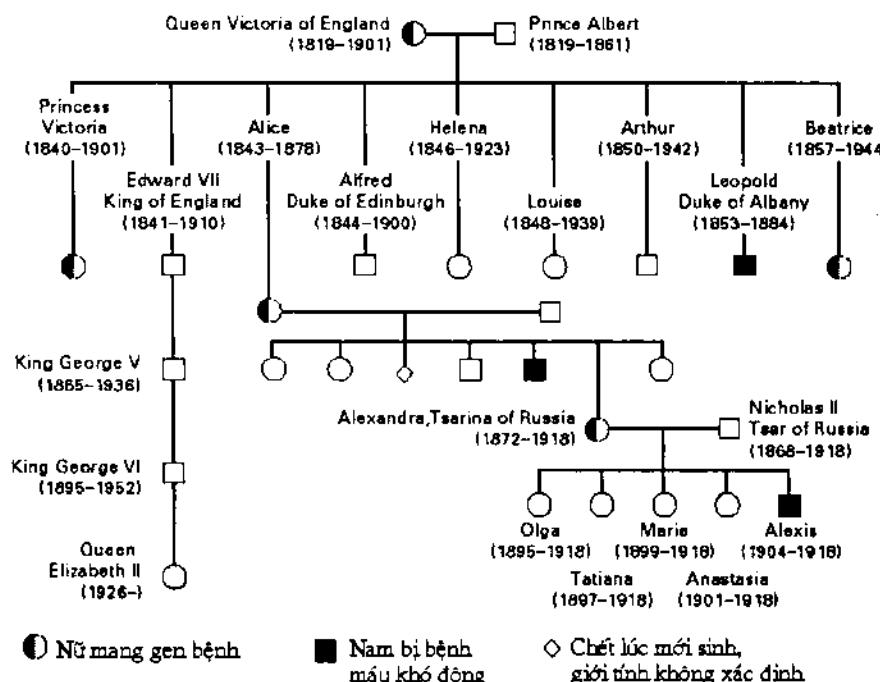
#### 4. Sản xuất yếu tố VIII và yếu tố IX tái tổ hợp

##### 4. 1. Lược sử nghiên cứu yếu tố VIII và yếu tố IX (Factor VIII & IX)

Bệnh máu khó đông, là một trong những loại bệnh do rối loạn di truyền gây nên, bệnh gặp chủ yếu ở nam giới với tần suất trung bình 10 000 người có một người bị bệnh. Bệnh máu khó đông rất hiếm gặp ở phụ nữ.

Từ năm 1110, đã phát hiện một số bệnh nhân nam bị chấn thương, không có khả năng cầm máu do máu khó đông, dẫn đến tử vong do mất máu.

Trường hợp di truyền bệnh máu khó đông điển hình được phát hiện và nghiên cứu ở gia đình Nữ Hoàng Anh Victoria (1819 -1901). Trong gia đình này, gen đột biến gây bệnh máu khó đông được di truyền cho nhiều thế hệ con cháu.



Hình 6. 18. Sự di truyền bệnh máu khó đông trong gia đình

Nữ Hoàng Anh (1819 -1901)

Năm 1828, bệnh máu khó đông được gọi là Hemophilia (bệnh ưa chảy máu). Năm 1936, Patek và Stetson đã phát hiện trong huyết tương của bệnh nhân Hemophilia thiếu chất gây đông máu. Năm 1952, phát hiện có 2 loại bệnh máu khó đông khác nhau gọi là Hemophilia A và Hemophilia B.

Năm 1959, khi nghiên cứu các yếu tố tham gia quá trình đông máu Wright đề nghị đánh số tất cả các chất ảnh hưởng đến sự đông máu từ I đến XII. Trong các yếu tố tham gia quá trình đông máu hai yếu tố quan trọng nhất là yếu tố VIII và yếu tố IX. Sự phối hợp của yếu tố VIII và yếu tố IX dẫn đến hình thành yếu tố X, sau đó tạo nên phức hệ enzym prothrombinase, dẫn đến sự hình thành thrombin tạo nên Fibrinogen và các sợi Fibrin gây đông máu. Sự thiếu hụt yếu tố VIII hoặc IX dẫn tới sự thiếu hụt các chất cần thiết cho quá trình đông máu. Bệnh Hemophilia A do những rối loạn di truyền liên quan đến yếu tố VIII, còn bệnh Hemophilia B liên quan đến rối loạn yếu tố IX.

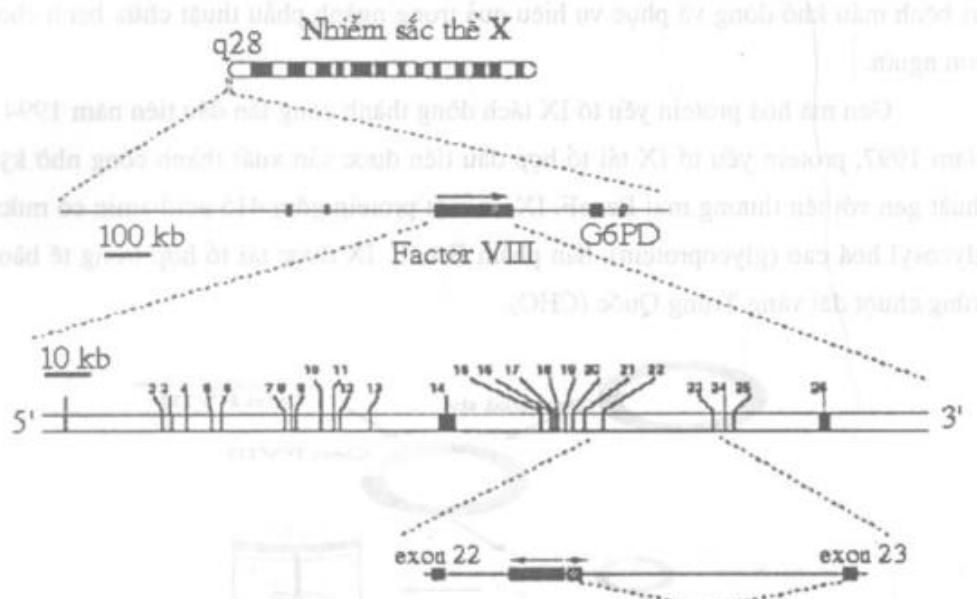
Năm 1984, Công ty Genetech tách dòng được gen mã hoá yếu tố VIII (Factor VIII). Năm 1992, yếu tố VIII tái tổ hợp được tổng hợp thành công. Năm 1994, yếu tố IX tái tổ hợp được tổng hợp.

#### 4. 2. Gen mã hoá yếu tố VIII và yếu tố IX

Protein yếu tố VIII được mã hoá bởi gen F.VIII nằm trên cánh dài của nhiễm sắc thể X, ở vị trí Xq28 cách vị trí telomer khoảng 1,1 Mb. Gen F.VIII có kích thước 186 kb. Trong gen F.VIII có 26 exon, các exon có kích thước nhỏ, hàm lượng DNA trong các exon chiếm khoảng 5% tổng hàm lượng DNA của gen. Gen F.VIII có 25 intron kích thước rất lớn, chiếm 95 % kích thước gen.

Bệnh Hemophilia A phần lớn liên quan đến các đột biến mất đoạn hoặc thêm đoạn trong gen F.VIII, một số nghiên cứu năm 2002 cho thấy có hơn 800 đột biến ở gen này (<http://europium.csc.mrc.ac.uk/usr/>). Bệnh Hemophilia A gây nên do thiếu hụt yếu tố VIII dẫn đến bệnh Hemophilia A. Yếu tố VIII (Factor VIII) là một glycoprotein trong huyết tương máu người, phân tử protein yếu tố VIII gồm 2 332 acid amin, trọng lượng phân tử khoảng 256 000 Dalton. Yếu tố VIII được tổng

hợp trong các tế bào gan, có thời gian tồn tại ngắn khoảng 8 -12 giờ, hàm lượng khoảng 100 - 200ng/ml máu.



Hình 6. 19. Sơ đồ cấu trúc gen yếu tố VIII (Factor VIII) với 26 exon

Bệnh Hemophilia B có tần xuất thấp hơn so với Hemophilia A, cứ khoảng 30 000 dân ông có một người bị bệnh. Phần lớn các bệnh Hemophilia B do rối loạn di truyền và đột biến ở gen yếu tố IX (F. IX). Gen F.IX nằm trên nhiễm sắc thể giới tính X, ở vị trí Xq27. Kích thước gen F.IX khoảng 34 kb, gồm 8 exon, mã hoá phân tử protein gồm 415 acid amin. Trên gen F. IX đã phát hiện 896 điểm đột biến, các đột biến trên gen F.IX đều gây ảnh hưởng đến mức độ đông máu của người.

(<http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html>, 2003)

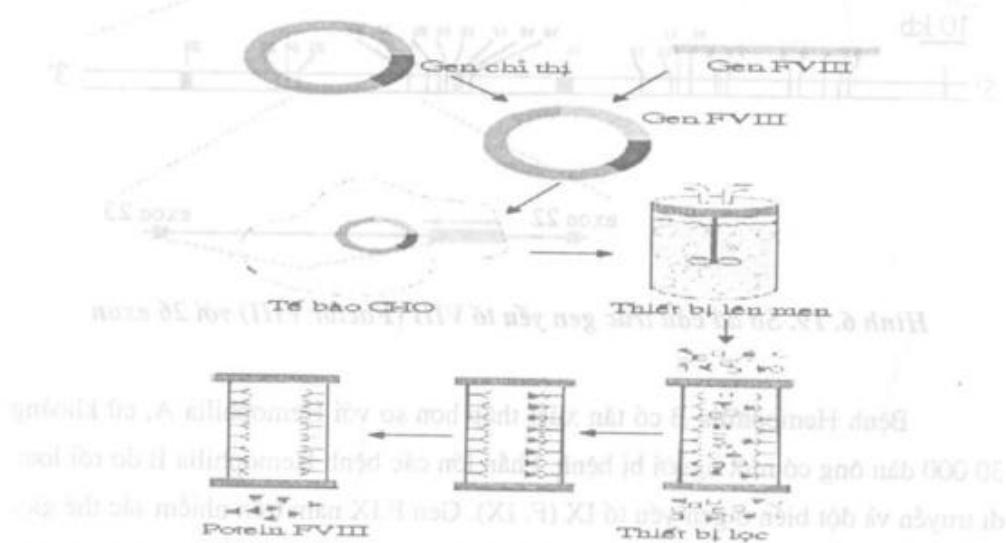
#### 4. 3. Công nghệ sản xuất chế phẩm yếu tố VIII, yếu tố IX tái tổ hợp

Trước đây, bằng các kỹ thuật tách chiết yếu tố VIII, và yếu tố IX từ huyết

tương máu người tạo nên các sản phẩm thương mại chữa bệnh Hemophilia. đã góp

phản diêu trị bệnh Hemophilia ở người, tuy nhiên hiệu quả chưa cao, giá thành đắt nên ít được ưa chuộng. Ngày nay bằng kỹ thuật tái tổ hợp, nhiều sản phẩm yếu tố VIII và yếu tố IX được thương mại hoá, và ứng dụng tương đối rộng rãi trong điều trị bệnh máu khó đông và phục vụ hiệu quả trong ngành phẫu thuật chữa bệnh cho con người.

Gen mã hoá protein yếu tố IX tách dòng thành công lần đầu tiên năm 1994. Năm 1997, protein yếu tố IX tái tổ hợp đầu tiên được sản xuất thành công nhờ kỹ thuật gen với tên thương mại BeneF. IX, là một protein gồm 415 acid amin có mức glycosyl hoá cao (glycoprotein). Sản phẩm BeneF. IX được tái tổ hợp trong tế bào trứng chuột đất vàng Trung Quốc (CHO).



Hình 6. 20. Sơ đồ công nghệ sản xuất yếu tố VIII và yếu tố IX tái tổ hợp

Hiện nay, bằng kỹ thuật gen tách dòng gen F.VIII từ tế bào người, thực hiện tái tổ hợp trong tế bào động vật có vú hoặc tế bào chuột đất vàng Trung Quốc (CHO - Chinese Hamster Ovary),..., người ta đã tạo được các yếu tố VIII thương mại khác nhau. Năm 1984, gen mã hoá yếu tố VIII ở người được tách dòng thành công.

Một số sản phẩm yếu tố VIII tái tổ hợp với tên thương mại Bioclote™ và

Recombinate™ được tái tổ hợp trong tế bào trứng chuột Hamster (CHO); Helixate và Kogenate được tái tổ hợp trong tế bào biểu mô ruột của chuột Hamster non, và nhiều sản phẩm khác ReFacto (2001), Advate (2003) ... góp phần điều trị bệnh Hemophilia A có hiệu quả cao.

Một số sản phẩm yếu tố VIII chủ yếu gồm: Alphanate, Humate-PR, Koate - HP, Hemophil ...

Một số sản phẩm yếu tố IX gồm: Bebulin™ VH, Konyne 80, Profilnine SD, Proplex T, AlphaNine ...

Công nghệ sản xuất yếu tố VIII được thực hiện với các bước cơ bản, giống như sản xuất các protein tái tổ hợp khác:

Tách chiết mRNA thông tin từ tế bào người, thực hiện kỹ thuật phiên mã ngược tạo cDNA mạch kép. Tạo vectơ tái tổ hợp, chuyển vào tế bào trứng của chuột đất vàng Trung Quốc, thực hiện quá trình tổng hợp protein tái tổ hợp trong các thiết bị lén men. Thu sinh khối, phá vỡ cấu trúc tế bào và tiến hành tách protein yếu tố VIII bằng các thiết bị lọc phân đoạn, sau đó sử dụng các kỹ thuật tinh sạch thu sản phẩm.

## 5. Sản xuất vaccin tái tổ hợp và vaccin DNA

### 5. 1. Lược sử nghiên cứu và ứng dụng vaccin

Bệnh đậu mùa và cách chữa bệnh đậu mùa có từ thời cổ xưa. Các nhà khảo cổ học đã phát hiện các tài liệu ghi chép về bệnh đậu mùa ở Ai Cập từ 3 000 năm trước Công nguyên, bệnh đậu mùa ở Trung Quốc khoảng hai nghìn năm trước Công nguyên cũng như nhiều tài liệu ghi lại dịch bệnh đậu mùa ở Thổ Nhĩ Kỳ khoảng 1 500 năm trước Công nguyên.

Năm 1780, Edward Jenner phát hiện phương pháp phòng bệnh đậu mùa, bằng chủng đậu với các virus gây bệnh đậu bò giảm độc lực.

Năm 1885, Lui. Pasteur phát minh vaccin chữa bệnh dại (Rabies vaccine).

dịch tả - *Diphtheria* và vaccin phòng bệnh uốn ván - *Tetanus* (1920); vaccin đậu mùa - *Pertussis* (1934); vaccin bại liệt - *Salk polio* (1955); vaccin phòng bệnh quai bị - *Mumps*, sởi - *measles* (1960); vaccin phòng bệnh cúm (1985), vaccin phòng chống bệnh viêm gan (1990)....

Vaccin được chia làm nhiều loại khác nhau, tuỳ thuộc nguồn gốc và bản chất của vaccin. Có thể chia vaccin thành hai nhóm chủ yếu là vaccin truyền thống và vaccin tái tổ hợp. Vaccin truyền thống (còn gọi là vaccin cổ điển) gồm vaccin là virus và vi khuẩn sống đã làm giảm độc, vaccin là virus và vi khuẩn chết lực (vaccin vô hoạt). Các loại vaccin truyền thống đã và đang giúp phòng chống nhiều loại bệnh truyền nhiễm cho người và động vật. Tuy nhiên, vaccin truyền thống có một số nhược điểm: thời gian bảo quản ngắn, khó đảm bảo an toàn tuyệt đối do vi khuẩn, virus dễ bị độc hoá trở lại...

Vaccin tái tổ hợp là các loại vaccin tạo từ các protein kháng nguyên được sản xuất nhờ kỹ thuật gen và các vaccin DNA tái tổ hợp. Vaccin tái tổ hợp có độ an toàn tuyệt đối, hiệu quả cao trong phòng chống nhiều loại bệnh truyền nhiễm, không gây hại với sức khoẻ con người, nên ngày càng được sử dụng rộng rãi. Vaccin tái tổ hợp được sản xuất ở qui mô công nghiệp có giá thành rẻ, được sử dụng rộng rãi trong phòng chống nhiều loại bệnh do nhiễm vi khuẩn (*Streptococcus pyogenes*, *Haemophilus influenza...*), bệnh do ký sinh trùng (bệnh sốt rét), bệnh do nhiễm virus: viêm gan B, A, C, D; virus *Epstein-Barr*; virus *ebola*; HIV.....

Vaccin DNA là loại vaccin có hiệu quả cao trong phòng chống nhiều bệnh hiểm nghèo (ung thư, HIV...). Vaccin DNA hiện đang ở giai đoạn thử nghiệm, do giá thành sản phẩm còn cao, chưa được ứng dụng rộng rãi. Vaccin DNA là loại vaccin mới, có nhiều triển vọng điều trị nhiều loại bệnh cho con người trong tương lai.

### **5. 2. Vaccin tái tổ hợp**

Vaccin tái tổ hợp là các loại vaccin sản xuất bằng kỹ thuật gen, nhờ tạo các sản phẩm protein kháng nguyên tái tổ hợp trong các tế bào chủ. Vaccin tái tổ hợp

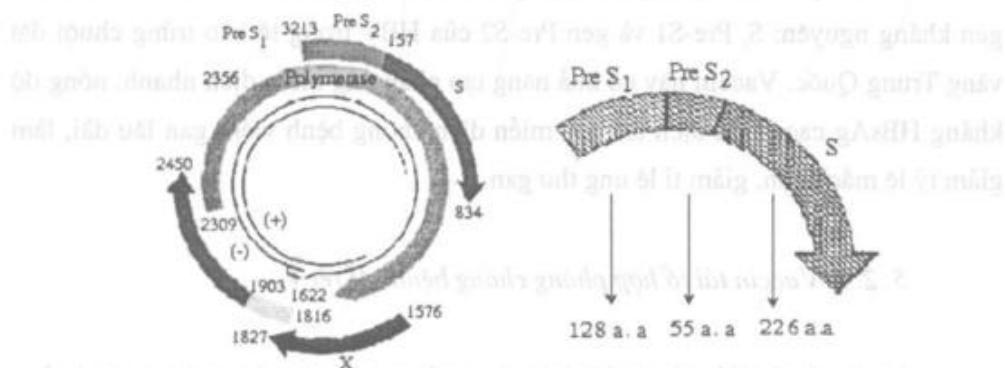
điển hình là vaccin phòng chống các bệnh viêm gan (B, A, C....), vaccin phòng chống HIV, và các bệnh sốt xuất huyết, sốt rét, cúm gà...

### 5.2.1. Vaccin tái tổ hợp phòng chống bệnh viêm gan

Bằng kỹ thuật gen, kháng nguyên bề mặt virus viêm gan B (HBV) được sản xuất ở nấm men *S. cerevisiae* từ năm 1981. Vaccin tái tổ hợp được sử dụng phòng chống bệnh viêm gan cho người bắt đầu từ 1986, đến nay đã có nhiều triệu người được tiêm chủng và điều trị bệnh viêm gan bằng vaccin tái tổ hợp.

Các sản phẩm thương mại vaccin tái tổ hợp phòng chống viêm gan B khác nhau: Recombivax HB, Engerix-B, Genhevac B ... được sản xuất từ các protein tái tổ hợp trong tế bào nấm men (*S. cerevisiae*, *P. pastoris*, *E. coli*...), được ứng dụng rộng rãi ở nhiều quốc gia.

Vaccin tái tổ hợp phòng chống viêm gan B được sản xuất trên cơ sở tách dòng các gen mã hoá protein vỏ của virus HBV (Hepatitis B Virus), tái tổ hợp và biểu hiện trong các tế bào chủ (nấm men, vi khuẩn...). Protein kháng nguyên HBsAg có khả năng tạo kháng thể miễn dịch mạnh nhất. Do đó, người ta thực hiện kỹ thuật tách dòng các gen mã hoá protein vỏ virus HBV, thực hiện tổng hợp protein tái tổ hợp, làm cơ sở sản xuất vaccin tái tổ hợp.



Hình 6.21. Cấu trúc bộ gen virus HBV và các loại protein vỏ

Virus HBV gen mã hoá protein vỏ gồm 3 gen: gen Pre - S1 mã hoá chuỗi

peptid có 128 acid amin, gen Pre - S2 mã hoá peptid có 55 acid amin và gen S mã hoá chuỗi peptid có 226 acid amin.

Bằng kỹ thuật PCR và các kỹ thuật phân tử khác, có thể tách riêng gen kháng nguyên S, gen mã hoá cho protein pre - S<sub>2</sub> được tách dòng năm 1981, được tái tổ hợp ở tế bào nấm men *Saccharomyces cerevisiae* (Rutter và cộng sự, Đại học California). Người ta có thể tách riêng phức hợp gen S với gen Pre - S<sub>1</sub>, hoặc phức hợp gen S với gen Pre - S<sub>2</sub> tạo nên các vectơ tái tổ hợp khác nhau. Biểu hiện các loại vectơ tái tổ hợp khác nhau thu các loại protein tái tổ hợp khác nhau, tạo nên các loại vaccine có hiệu quả khác nhau.

Vaccine viêm gan B tái tổ hợp biểu hiện trong nấm men và thương mại hoá đầu tiên 1986 với tên thương mại Engerix B (Smith Kline, Mỹ). Hiện nay có nhiều sản phẩm thương mại khác nhau được sử dụng rộng rãi: Recombivax (Merck-Sharp, Mỹ); Genhevac B (Pasteur Meraux)....

Việt Nam là một nước có mức nhiễm HBV trong cộng đồng dân cư ở mức cao: tỉ lệ viêm gan cấp tính do HBV khoảng 37 - 49%, tỉ lệ xơ gan có HBV là 47 - 87% và tỷ lệ ung thư gan có HBV là 57-80%. Hiện nay, Việt Nam là một trong những nước sản xuất thành công vaccine viêm gan B và vaccine viêm gan A tái tổ hợp trong tế bào nấm men *Pichia pastoris* ở qui mô công nghiệp có hiệu quả cao.

Gần đây, một loại vaccine viêm gan B mới được tạo nên từ sự biểu hiện cả gen kháng nguyên: S, Pre-S<sub>1</sub> và gen Pre-S<sub>2</sub> của HBV trong tế bào trứng chuột đất vàng Trung Quốc. Vaccine này có khả năng tạo phản ứng miễn dịch nhanh, nồng độ kháng HBsAg cao miễn dịch tốt, tạo miễn dịch chống bệnh viêm gan lâu dài, làm giảm tỷ lệ mắc bệnh, giảm tỉ lệ ung thư gan.

### *5. 2. 2. Vaccine tái tổ hợp phòng chống bệnh sốt rét*

Bệnh sốt rét chủ yếu do ký sinh trùng sốt rét gây nên, hàng năm có khoảng 300 - 500 triệu người nhiễm bệnh, gây tử vong từ 1 - 3 triệu người. Ký sinh trùng sốt rét có 4 loài chủ yếu, có vòng đời và quá trình lây nhiễm rất phức tạp, trong mỗi giai đoạn phát triển của ký sinh trùng sốt rét cần các loại thuốc đặc trị khác nhau.

Hiện nay, vaccine phòng chống sốt rét chủ yếu cắt một giai đoạn trong chu trình phát triển của ký sinh trùng. Trong các loài ký sinh trùng, loài *Plasmodium falciparum* gây bệnh phổ biến nhất, do vật trung gian là muỗi truyền bệnh. Do vậy, vaccine phòng chống sốt rét tái tổ hợp từ một gen kháng nguyên đặc hiệu *PfS* - 25, là vaccine phòng chống bệnh sốt rét đang được thử nghiệm. Gen *PfS* - 25 phân lập từ loài *Plasmodium falciparum* mã hoá một protein có trọng lượng phân tử 25 kD, được đưa vào vectơ tái tổ hợp và biểu hiện thành công trong tế bào nấm men. Peptid tái tổ hợp tạo nên có khả năng kích thích tạo kháng thể mạnh ở các tế bào thử nghiệm, triển vọng là vaccine phòng chống sốt rét tốt trong những năm tới.

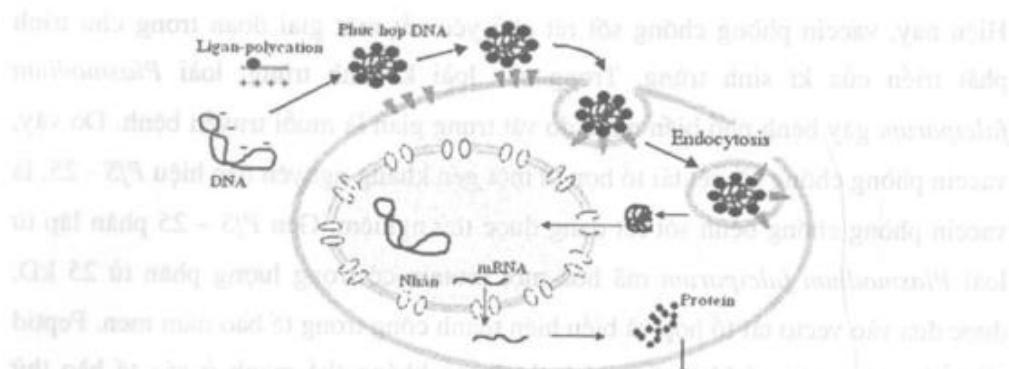
Hiện nay, nhiều loại vaccine tái tổ hợp phòng chống bệnh viêm não, bệnh cúm gà... đã được sử dụng rộng rãi, có hiệu quả phòng chống bệnh cao. Các loại vaccine phòng chống bệnh sốt rét, bệnh sốt xuất huyết, phòng chống HIV/AIDS đang được nghiên cứu và sản xuất thử nghiệm ở một số nước có kết quả.

### 5. 3. Vaccine DNA

#### 5. 3. 1. *Lược sử nghiên cứu và ý vai trò của vaccine DNA*

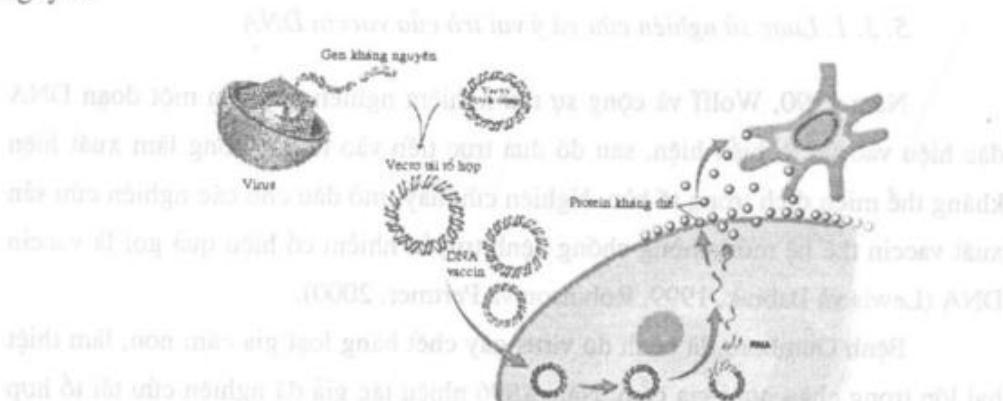
Năm 1990, Wolff và cộng sự thử nghiệm nghiên cứu gắn một đoạn DNA đặc hiệu vào vectơ biểu hiện, sau đó đưa trực tiếp vào tế bào sống làm xuất hiện kháng thể miễn dịch trong tế bào. Nghiên cứu này, mở đầu cho các nghiên cứu sản xuất vaccine thế hệ mới, phòng chống bệnh truyền nhiễm có hiệu quả gọi là vaccine DNA (Lewis và Babiuk, 1999; Robinson và Pertmer, 2000).

Bệnh Gumboro, là bệnh do virus gây chết hàng loạt gia cầm non, làm thiệt hại lớn trong chăn nuôi gia cầm. Năm 1996 nhiều tác giả đã nghiên cứu tái tổ hợp đơn gen kháng nguyên VP2 hoặc các đoạn gen VP3, VP4 tạo vaccine DNA, có hiệu quả cao trong phòng chống bệnh Gumboro cho gia cầm (Fodor và cộng sự, 1999). Vaccine DNA phòng bệnh Gumboro đã đăng ký bản quyền ở Mỹ, và được sử dụng ở nhiều nước trên thế giới (Aboud - Pirak và cộng sự, 2002). Gần đây, vaccine DNA phòng chống HIV/AIDS, phòng chống ung thư và nhiều loại bệnh khác được sản xuất và thử nghiệm ở nhiều quốc gia.



**Hình 6. 22. Vaccin DNA là DNA trán được bao gói bởi polycation**

Vaccin DNA gồm 2 loại chủ yếu gồm loại vaccin tập hợp các gen kháng nguyên và loại tập hợp các gen sinh kháng thể trực tiếp. Gen kháng nguyên được tách dòng từ virus, vi khuẩn gây bệnh có thể ở dạng DNA trán được bao gói bởi polysom, polycation, màng lipid... hoặc là các vectơ tái tổ hợp mang gen kháng nguyên.



**Hình 6. 23. Vaccin DNA là vectơ tái tổ hợp**

Trong tế bào, các gen kháng nguyên được phiên mã và dịch mã tạo nên các protein kháng nguyên, các protein kháng nguyên kích thích tế bào sản xuất kháng thể đặc hiệu chống các tác nhân gây bệnh là virus, vi khuẩn...Vaccin DNA là tập

hợp các gen mã hoá kháng thể đặc hiệu thường được cài gắn vào các vectơ biểu hiện mạnh, khi đưa vào tế bào các gen được dịch mã tạo các kháng thể đặc hiệu, giúp tế bào chống lại sự xâm nhiễm của các tác nhân gây bệnh.

Vaccin DNA không chỉ có hiệu quả đáp ứng miễn dịch cao, giúp phòng chống bệnh có hiệu quả, còn có nhiều ưu điểm: dễ dàng thiết kế và sản xuất, bảo quản được lâu, vận chuyển và sử dụng đơn giản và tiện lợi....

Vaccin DNA có thể bảo quản hàng năm ở - 20 °C vẫn giữ được hoạt tính, ở điều kiện nhiệt độ thường có thể giữ được trong 1- 2 tháng, do vậy không cần bảo quản lạnh khi đưa vaccin đi sử dụng. Một khác, vaccin DNA rất dễ sử dụng: vaccin DNA tinh khiết được pha loãng trong môi trường nước muối sinh lý (NaCl 0,9%) hoặc PBS (phosphate buffer saline) và tiêm vào cơ bắp một lượng nhỏ (0, 2 - 0,5 ml). Ngoài ra, có thể sử dụng phương tiện cơ học khác là *súng bắn gen* (gene gun) để đưa vaccin DNA vào cơ thể, các phương pháp này có thể đưa vaccin DNA trực tiếp đến mô, tế bào bị bệnh hoặc tiêm trực tiếp vaccin DNA chống ung thư vào giữa khối u làm tăng nhanh hiệu quả tác dụng của vaccin. Lưu ý, cùng một loại vaccin DNA khi sử dụng các phương pháp đưa vaccin vào cơ thể khác nhau, hiệu quả của vaccin khác nhau do ảnh hưởng của cơ chế đáp ứng miễn dịch tế bào.

Một đặc điểm rất thuận lợi của vaccin DNA là dễ dàng tạo được vaccin DNA đa chủng và vaccin DNA đa giá, giúp cơ thể chống được nhiều loại bệnh nhiễm trùng với giá thành thấp. Vaccin DNA đa giá được sản xuất bằng cách thiết kế một vectơ tái tổ hợp mang nhiều gen kháng nguyên kháng nhau, còn vaccin DNA đa chủng được tạo bằng cách trộn nhiều loại vectơ tái tổ hợp mang gen kháng nguyên khác nhau trước khi đưa vào cơ thể.

Với nhiều điểm ưu việt, giá thành rẻ, hiệu quả chống bệnh cao, nên vaccin DNA ngày càng được ưa chuộng và sử dụng. Trong những năm gần đây có nhiều loại vaccin DNA chống virus, vi khuẩn và ký sinh trùng được sản xuất, sử dụng trong phòng chống bệnh nhiễm trùng cho con người và vật nuôi.

Hiện nay, có một số loại vaccin DNA chống các bệnh truyền nhiễm cho người được đưa vào sử dụng ở mức thử nghiệm. Vaccin DNA chống ung thư được

thử nghiệm vào thập kỉ 90 do các nhà khoa học ở Southampton và Cambridge sản xuất. Sau khi tiêm vaccin DNA chống ung thư, trong tế bào hình thành kháng thể chống ung thư làm giảm sự phân chia của các tế bào ung thư. Vaccin DNA chống lao được thiết kế từ DNA của vi khuẩn lao và thử nghiệm trên chuột nhắt năm 1986, các nhà khoa học Anh tạo nên những con chuột nhắt miễn dịch với bệnh lao. Vaccin DNA thiết kế từ bộ gen virus HIV- 1 đã được thử nghiệm phòng chống HIV-1 trên chuột có hiệu quả cao. Chương trình thử nghiệm vaccin DNA phòng chống HIV/ AIDS cho người tình nguyện đang được thực hiện có kết quả ở một số nước Châu Phi, Châu Á.... Một số nghiên cứu vaccin DNA phòng chống bệnh cúm gà H5N1, đã được nghiên cứu và thử nghiệm ở nhiều nước cho kết quả tốt. Vaccin DNA là một trong những loại vaccin có triển vọng rất lớn trong phòng chống bệnh truyền nhiễm, bảo vệ sức khoẻ con người.

Cơ chế tạo miễn dịch khi sử dụng vaccin DNA đã được hiểu biết đầy đủ. Tuy nhiên, người ta vẫn e ngại vấn đề an toàn của vaccin DNA. Một vấn gen kháng nguyên (từ virus, vi khuẩn) có xâm nhập vào (integration) trong hệ gen người và động vật hay không? là câu hỏi luôn luôn được đặt ra. Hiện nay, chưa có bằng chứng nào về sự xâm nhập gen có trong vaccin DNA vào cơ thể người, tuy nhiên cần có các nghiên cứu thêm để khẳng định tính an toàn của vaccin DNA (Babiuk và cộng sự, 2002). Do đó, vaccin DNA phòng chống các bệnh truyền nhiễm, ung thư ở người mới được phép thử nghiệm trên người tình nguyện hoặc ở động vật có vú (Griffiths, 1995; Robinson và Pertmer, 2000). Vaccin DNA được WHO cho phép sử dụng rộng rãi trong các chương trình vaccin phòng chống bệnh ở động vật (Watts và Kennedy, 1999; Babiuk và cộng sự, 2002).

Trong sản xuất vaccin DNA, cần thiết kế các vectơ tái tổ hợp mang một promoter mạnh, đảm bảo cho gen protein kháng nguyên (hoặc gen mã hoá protein kháng thể) được biểu hiện, một số promoter mạnh của virus thường được sử dụng để biểu hiện các gen kháng nguyên như: promoter CMV (cytomegalovirus), promoter SV 40 (simian virus), promoter RSV (rous sarcoma virus), TK promoter (gen thymidine kinase)... (Robinson và Pertmer, 2000).

### 5. 3. 2. Ví dụ về công nghệ sản xuất vaccin DNA phòng chống bệnh Gumboro

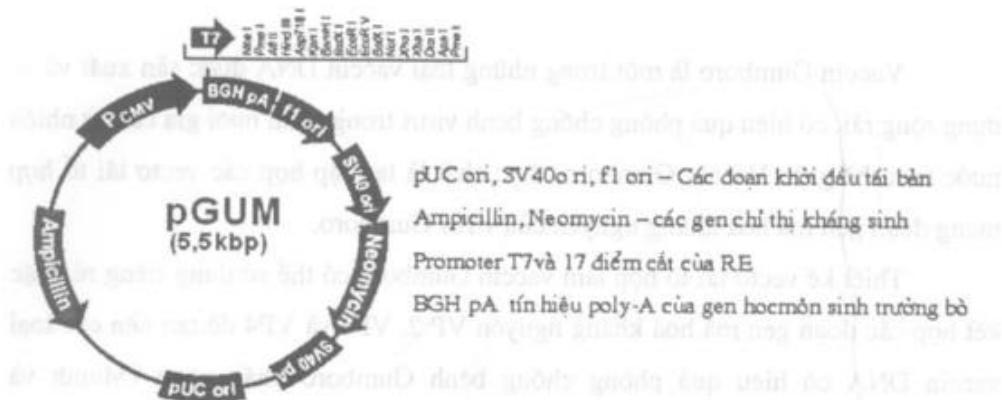
Vaccin Gumboro là một trong những loại vaccin DNA được sản xuất và sử dụng rộng rãi, có hiệu quả phòng chống bệnh virus trong chăn nuôi gia cầm ở nhiều nước trên thế giới. Vaccin Gumboro, thực chất là tạo tập hợp các vectơ tái tổ hợp mang đoạn gen mã hoá kháng nguyên của virus Gumboro.

Thiết kế vectơ tái tổ hợp làm vaccin Gumboro có thể sử dụng riêng rẽ hoặc kết hợp các đoạn gen mã hoá kháng nguyên VP 2, VP3 và VP4 để tạo nên các loại vaccin DNA có hiệu quả phòng chống bệnh Gumboro khác nhau (Mundt và Vakharia, 1996; Chang và cộng sự, 2002).

Vaccin DNA chứa 2 đoạn gen mã hoá protein vỏ virus Gumboro gồm đoạn gen VP2 và phân đoạn A mang cả 3 gen (VP2, VP4 và VP3), tách dòng từ các chủng virus Gumboro P40 và D78 sử dụng phòng chống bệnh cho gia cầm ở Hungary có hiệu quả cao (Fodor và cs, 1999). Vaccin DNA chứa phân đoạn A gồm 3 gen (VP2, VP4 và VP3) từ chủng Gumboro STC, sử dụng trong phòng chống virus Gumboro thành công tại Mỹ và nhiều nước khác. Vaccin DNA phòng bệnh Gumboro được sử dụng ở nhiều nước trên thế giới, đã được đăng ký bản quyền chính thức ở Mỹ (Aboud-Pirak và cộng sự, 2002).

Vaccin DNA kích thích sinh kháng thể trung hoà, đồng thời tạo miễn dịch trung gian tế bào, nên có tác dụng miễn dịch tương đối lâu bền (Chang và cộng sự, 2002).

Sản xuất vaccin Gumboro có thể sử dụng nhiều loại vectơ biểu hiện gen khác nhau, trong đó vectơ pGUM với promoter mạnh CMV (cytomegalovirus) được sử dụng khá phổ biến. Nguồn cung cấp gen mã hoá kháng nguyên được tách chiết từ nhiều chủng Gumboro đã chọn lọc như: chủng virus Gumboro STC, chủng GLS, chủng D78 .... (Aboud-Pirak và cộng sự, 2002). Các đoạn gen mã hoá protein vỏ virus Gumboro chủ yếu là VP 2, VP3 và VP4 được sử dụng làm gen kháng nguyên.



Hình 6. 24. Vector pGUM trong sản xuất vaccin DNA Gumboro

Qui trình công nghệ sản xuất vaccin DNA chống bệnh Gumboro có thể thực hiện theo các bước cơ bản sau:

Tách chiết RNA tổng số từ mẫu virus Gumboro (sử dụng Kit tách chiết RNA thông dụng).

Thực hiện phản ứng RT-PCR với các cặp mồi đặc hiệu để thu nhận đoạn gen mã hoá protein kháng nguyên cần thiết. Các cặp mồi đặc hiệu có thể mua từ hàng sinh phẩm, hoặc thiết kế theo điều kiện cụ thể của phòng thí nghiệm, theo nguyên tắc thiết kế mồi thông dụng.

Ví dụ, để nhân phứ hợp 3 gen VP 2, VP3, VP 4 có thể sử dụng cặp mồi:  
 Mồi xuôi (F): 5' - CCGGAATTCCAAACGATCGCAGCGATGACAAAC - 3'  
 Mồi ngược (R): 5' - ATAAGAATGCGGCCGCTTATCACAAAGGTCCATC  
 AGAGACA - 3'

Tạo vector tái tổ hợp giữa đoạn gen kháng nguyên với vectơ tách dòng. Thực hiện kỹ thuật biến nạp, sàng lọc các tế bào mang vectơ tái tổ hợp. Nuôi cấy các tế bào sau biến nạp, thu nhận và kiểm tra DNA tái tổ hợp, thử hiệu lực vacxin DNA.

## II. Sử dụng kỹ thuật gen trong chẩn đoán và chữa bệnh

### 1. Sử dụng kỹ thuật gen trong chẩn đoán và điều trị ung thư do virus

#### 1. 1. Ứng dụng kỹ thuật gen trong chẩn đoán ung thư do virus

Kỹ thuật PCR là một trong những kỹ thuật được sử dụng để phát hiện nhanh các gen đặc hiệu, gen kháng nguyên của nhiều loại virus gây ung thư, giúp chẩn đoán sớm bệnh nhân ung thư. Sử dụng kỹ thuật gen trong chẩn đoán sớm, chẩn đoán nhanh bệnh ung thư đã được ứng dụng ở nhiều bệnh viện tại Mỹ, Anh, Đức, Thái Lan, Nhật, Đài Loan.... Các bệnh ung thư vòm họng, ung thư gan, ung thư dạ dày..... do virus đã được chẩn đoán nhanh, chẩn đoán sớm bằng kỹ thuật PCR hoặc kỹ thuật lai phân tử đánh dấu huỳnh quang.

Ung thư vòm họng phần lớn do virus EBV (*Epstein-Barr virus*) gây nên. *Epstein-Barr virus* thuộc họ virus Herpes, có bộ gen là phân tử DNA mạch kép, kích thước bộ gen khoảng 184 kb. Bằng kỹ thuật PCR, với các cặp mồi đặc hiệu khác nhau nhận các đoạn gen đặc trưng của virus EBV, có thể chẩn đoán ung thư vòm họng chính xác trên 90% (Saito và cộng sự, 1989).

Hiện nay đã phát hiện có 6 loại virus gây viêm gan nguyên phát (HAV, HBV, HCV, HDBV, HEV, HGV). Trong đó, virus HBV (Hepatitis B virus) có bộ gen DNA, còn các loại virus viêm gan khác có bộ gen RNA. HBV là tác nhân gây viêm gan dẫn đến ung thư gan nhiều nhất, ung thư gan do HBV chiếm khoảng 30% các bệnh ung thư có liên quan với virus (Đái Duy Ban và cộng sự, 2002). Có nhiều phương pháp chẩn đoán ung thư gan khác nhau: chẩn đoán lâm sàng, chẩn đoán huyết thanh, chẩn đoán hình ảnh, chẩn đoán bằng kỹ thuật phân tử... trong đó chẩn đoán bằng kỹ thuật PCR cho kết quả nhanh với độ chính xác cao, đã được ứng dụng rộng rãi ở nhiều nước.

Ở Việt Nam, các nhà khoa học đã xây dựng được một số qui trình chẩn đoán nhanh các bệnh ung thư vòm họng, ung thư gan... bằng kỹ thuật PCR. Sứ

dụng kỹ thuật PCR với cặp mồi đặc hiệu TH<sub>1</sub> và TH<sub>2</sub> có thể chẩn đoán ung thư vòm họng bằng kỹ thuật PCR. Trình tự cặp mồi TH<sub>1</sub>, TH<sub>2</sub>:

Mồi TH<sub>1</sub>: 5' - AGCCAATTGTCAGTTCTAGGGAGGG - 3'

Mồi TH<sub>2</sub>: 5' - GCTTGGATGGCCGAGTCAGCGACGG - 3'

Sử dụng kỹ thuật PCR lồng (Nested PCR) với 2 cặp mồi HBV P<sub>1</sub> và HBV P<sub>2</sub>, các tác giả thuộc Viện Khoa học và Công nghệ Việt Nam đã phân lập được gen mã hoá kháng nguyên HBsAg của virus HBV. Hai cặp mồi sử dụng trong kỹ thuật PCR lồng để chẩn đoán sớm ung thư gan do virus gồm:

Cặp mồi 1:

Mồi dương (HBV P<sub>1</sub>): 5' - AGTAAACCCCTGTTCCGACTACTGCC - 3'

Mồi âm (HBV M<sub>1</sub>): 5' - AGTTGGCGAGAAAGTGAAGCCTG - 3'

Cặp mồi 2:

Mồi dương (HBV P<sub>2</sub>): 5' - GAACATGGAGAGCACAAACATCCAGG - 3'

Mồi âm (HBV M<sub>2</sub>): 5' - GCATATAAAGGCATCAAAGGCAGG - 3'

Tách chiết DNA từ các tế bào gan người bệnh, thực hiện kỹ thuật PCR lồng, có thể chẩn đoán ung thư gan nhanh bằng kỹ thuật PCR (Đái Duy Ban và cộng sự, 2004).

### **1. 2. Kỹ thuật gen và các ứng dụng trong điều trị bệnh ung thư**

Bệnh ung thư là loại bệnh có tỉ lệ gây tử vong cao nhất, có thể áp dụng nhiều biện pháp khác nhau điều trị ung thư: hoá chất, trị xạ, thuốc ức chế khối u... Một trong những phương pháp điều trị ung thư có hiệu quả là điều trị bằng liệu pháp gen. Phương pháp phổ biến trong điều trị ung thư ở người là sử dụng vectơ liệu pháp gen từ *paramyxovirus* cài biến gen, kết hợp trị xạ. Vectơ liệu pháp gen được đưa vào bệnh nhân bị ung thư bằng vi tiêm trực tiếp vào giữa khối u, tiêm vào tĩnh mạch hoặc đưa qua đường hậu môn....

Thực nghiệm liệu pháp gen điều trị ung thư cho một nhóm bệnh nhân ở Mỹ với 7 loại ung thư khác nhau (ung thư phổi (Lung), ung thư dạ dày (Gastric), ung thư tử cung (Uterine), ung thư da (Skin), ung thư gan (Hepatic), ung thư trực tràng

(Rectal), ung thư thực quản (Esophageal) thu được hiệu quả cao. Thí nghiệm điều trị 90 bệnh nhân, có 41 % bệnh nhân kích thước khối u giảm trên 50% và một số bệnh nhân có khối u biến mất hoàn toàn.

Các kết quả nghiên cứu thực nghiệm cho thấy, hơn 50% các trường hợp ung thư ở người liên quan đến mất hoạt tính gen *p53*. Gen *p53* ở người có kích thước 23 kb, nằm trên nhiễm sắc thể số 17. Gen *p53* mã hoá một protein có trọng lượng phân tử 53 kD (gọi là protein *p53*), protein *p53* có chức năng điều hòa quá trình phân chia tế bào và kích hoạt hiện tượng chết theo chu trình (apoptosis) của tế bào. Khi gen *p53* bị đột biến, mất khả năng kiểm soát sự phân chia tế bào, tế bào phân chia vô hạn hình thành các khối u. Sử dụng vector adenovirus mang gen *p53* (*Ad-p53*) chuyển vào người bệnh, có thể chữa khỏi hoặc làm giảm sự phát triển của khối u ung thư. Thực nghiệm tiêm trực tiếp vectơ *Ad-p53* vào khối u phế quản, các khối u nhỏ lại, tế bào ung thư ngừng phát triển. Nhiều loại ung thư được điều trị bằng liệu pháp gen *p53* có hiệu quả gồm ung thư phế quản, ung thư phổi, ung thư buồng trứng (ovarian cancer), ung thư đầu và cổ (head - neck), ung thư sắc tố (melanoma)....

Hiện nay, khoảng 60% các trường hợp bị bệnh ung thư được điều trị bằng sử dụng gen *p53*. Điều trị ung thư bằng liệu pháp gen sử dụng gen *p53*, kết hợp thêm một phần trị xạ hoặc sử dụng hoá chất là một trong những hướng điều trị bệnh ung thư có hiệu quả cao, được sử dụng ở nhiều bệnh viện trên thế giới.

## 2. Chẩn đoán một số bệnh nhiễm trùng bằng kỹ thuật PCR

Chẩn đoán sớm bệnh nhiễm virus, vi khuẩn ở người giúp các bác sĩ tìm ra phương pháp điều trị thích hợp, hiệu quả chữa bệnh cao. Ở các nước tiên tiến, trong nhiều bệnh viện đã áp dụng kỹ thuật phân tử để chẩn đoán nhanh và chính xác các bệnh viêm màng não, lao, sốt rét, sốt xuất huyết... Ở nước ta, áp dụng kỹ thuật PCR và các kỹ thuật phân tử trong chẩn đoán sớm bệnh truyền nhiễm còn nhiều hạn chế,

mới chỉ ở mức các nghiên cứu trong bệnh viện, viện nghiên cứu hoặc áp dụng thử nghiệm.

### **2. 1. Chẩn đoán sớm nhiễm ký sinh trùng sốt rét bằng kỹ thuật PCR**

Ở nước ta, trong những năm gần đây, Viện Sốt rét - Ký sinh trùng - Côn trùng Trung ương đã áp dụng kỹ thuật PCR lồng để chẩn đoán sớm sốt rét do ký sinh trùng gây bệnh. Kết quả chẩn đoán nhanh với độ chính xác rất cao, góp phần thực hiện chương trình quốc gia phòng chống bệnh sốt rét có hiệu quả.

Bệnh sốt rét ở người chủ yếu do 4 loài ký sinh trùng gây nên gồm: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale* và *Plasmodium malariae*. Trên 80% trường hợp sốt rét ở nước ta do nhiễm ký sinh trùng sốt rét thuộc loài *Plasmodium falciparum*, một số bệnh nhân nhiễm phôi hợp 2 - 3 loài ký sinh trùng sốt rét khác nhau. Chẩn đoán sớm và chính xác các loại ký sinh trùng ở bệnh nhân, giúp các bác sĩ xác định được các phác đồ điều trị thích hợp, công tác phòng chống sốt rét có hiệu quả cao. Sử dụng các cặp mồi Plu5 - Plu 6, Fal1 - Fal2, Viv1 - Viv2 và Mal1 - Mal2 có thể chẩn đoán nhanh chính xác loài ký sinh trùng lây nhiễm ở bệnh nhân.

Trình tự các cặp mồi gồm:

Cặp mồi PCR lần 1(Nested 1):

Plu5: 5' - CCTGTTGTTGCCTAAACTTC -3'

Plu6: 5' - TTAAAATTGTTGCAGTAAAGC -3'

Các cặp mồi PCR lần 2 (Nested 2), để xác định loài ký sinh trùng:

Cặp mồi xác định *Plasmodium falciparum*

Fal1: 5' - TTAAACTGGTTGGGAAAACCAAATATATT -3'

Fal2: 5' - ACACAATGAACCTAACATGACTACCCGTC -3'

Cặp mồi xác định *Plasmodium vivax*

Viv1: 5' - GCGTTCTAGCTTAATCCACATAACTGATAC -3'

Viv2: 5' - ACTTCCAAGCCGAAGCAAAGAAAGTCCTTA -3'

**Cặp mồi xác định *Plasmodium ovale***

Ova1: 5' - ATCTCTTTGCTATTTTAGTATTGGAGA -3'

Ova2: 5' - GGAAAAGGACACATTAATTGTATCCTAGTG -3'

**Cặp mồi xác định *Plasmodium malariae***

Mal1: 5' – ATAACATAGTTGTACGTTAAGAATAACCGC -3'

Mal2: 5' – AAAATTCCCATGCATAAAAAATTATACAAA -3'

Sử dụng kỹ thuật PCR lồng với 5 cặp mồi đặc hiệu, có thể chẩn đoán sớm tình trạng nhiễm ký sinh trùng sốt rét, cho phép xác định chính xác loài ký sinh trùng bệnh nhân bị nhiễm (Lê Định Công và cộng sự, 1996; Lê Đức Đào và cộng sự, 2003).

**2. 2. Chẩn đoán nhiễm vi khuẩn lao *M. tuberculosis* bằng kỹ thuật PCR**

Bệnh lao ở người có nhiều thể khác nhau: lao phổi, lao xương khớp, lao da, lao màng não... nguyên nhân chủ yếu do nhiễm vi khuẩn lao Mycobacterie tuberculosis. Chẩn đoán người nhiễm vi khuẩn lao bằng nhiều phương pháp khác nhau: chẩn đoán lâm sàng, chẩn đoán bằng phản ứng *Tuberculin*, chẩn đoán bằng vi khuẩn lao chết (BCG tes), chẩn đoán bằng kỹ thuật phân tử ELISA, PCR... Các kỹ thuật phân tử có độ nhạy cao, chuẩn đoán chính xác ngay khi mới chớm bệnh, giúp điều trị có hiệu quả.

Phương pháp PCR được áp dụng trong chẩn đoán bệnh lao ở nhiều nước trên thế giới, đạt độ chính xác 88% - 97% (Kox và cộng sự, 1996). Người ta thiết kế cặp mồi đặc hiệu, để nhận đoạn IS6110 là đoạn DNA đặc trưng của vi khuẩn lao *M. Tuberculosis*. Ở Việt Nam, một số ứng dụng kỹ thuật PCR chẩn đoán lao màng não ở người từ mẫu bệnh phẩm là dịch não tuỷ, kết quả chẩn đoán đạt 72,7 % (Trần Văn Sáng, 2000).

Ngoài ra, bệnh sốt xuất huyết do virus Dengue, và nhiều bệnh do rối loạn di truyền ở người (bệnh suy giảm miễn dịch ADA, bệnh run rẩy Parkinson...) cũng được chẩn đoán sớm bằng kỹ thuật gen có hiệu quả cao.

### *3. Sử dụng kỹ thuật gen trong chẩn đoán và điều trị bệnh máu khó đông (Hemophilia) ở người*

Bệnh máu khó đông là bệnh gây nên do rối loạn di truyền đơn gen. Có hai loại bệnh máu khó đông là hemophilia A và hemophilia B. Bệnh hemophilia A do bị đột biến, gây mất chức năng gen mã hoá một loại protein đặc hiệu gọi là yếu tố VIII (factor VIII, kí hiệu là F.VIII), bệnh chủ yếu ở đàn ông. Tần số mắc bệnh hemophilia A trung bình 10 000 người đàn ông có một người bị bệnh. Bệnh hemophilia B gây nên do rối loạn chức năng gen mã hoá yếu tố IX (factor IX, kí hiệu là F. IX), tần số mắc bệnh khoảng 1 trên 50 000 người.

Gen F.VIII nằm trên nhiễm sắc thể giới tính X của người, mã hoá protein yếu tố VIII có trọng lượng phân tử khoảng 300 kDaL. Gen F. IX, cũng nằm trên nhiễm sắc thể X của người, mã hoá protein yếu tố IX có trọng lượng phân tử khoảng 55 kDaL. Ở người bình thường hàm lượng protein yếu tố VIII trong máu khoảng 100 - 200 ng / ml máu, khi hàm lượng dưới 10 ng / ml máu gây hiện tượng máu khoá đông hoặc không đông.

Có thể sử dụng nhiều kỹ thuật phân tử khác nhau để phát hiện nhanh bệnh máu khó đông, xác lập phác đồ điều trị hữu hiệu. Hiện nay, để phát hiện nhanh bệnh hemophilia ở người có thể sử dụng các phương pháp PCR, RFLP cùng với các enzym giới hạn đặc hiệu. Để phát hiện các đột biến trên gen F.VIII, có thể sử dụng các enzym giới hạn khác nhau: enzym *Bcl* I phát hiện đột biến ở intron 18, *Hind* III phát hiện đột biến trên intron 19 hoặc các enzym *Xba* I và enzym *Msp* I phát hiện đột biến ở intron 22. Để phát hiện đột biến trên gen F. IX có thể sử dụng các enzym *Taq* I phát hiện đột biến ở intron 4 và enzym *Xmn* I phát hiện đột biến ở intron 3....

Ở nước ta, kỹ thuật PCR lồng được sử dụng trong kiểm tra và chẩn đoán bệnh nhân máu bệnh máu khó đông hemophilia, đã được thực hiện ở Viện Huyết học và Viện Vệ sinh dịch tễ Trung ương (Nguyễn Hạnh Phúc, 2003). Với các cặp mồi 8. 1 và 8. 2 hoặc C3 và C4 có thể phát hiện nhanh bệnh nhân bị máu khó đông bằng kỹ thuật PCR. Trình tự cặp mồi 8. 1 và 8. 2 sử dụng nhân đoạn intron 18 là:

Môи 8. 1: 5' - TAAAAGCTTAAATGGTCTAGGC -3'

Môи 8. 2: 5' - TTCAATTCTGAAATTATCTTTGT -3'

Cặp môи C3 và C4 sử dụng nhân đoạn đoạn intron 19:

Môи C3: 5' – TTCCCGAGCATCTACATGCT -3'

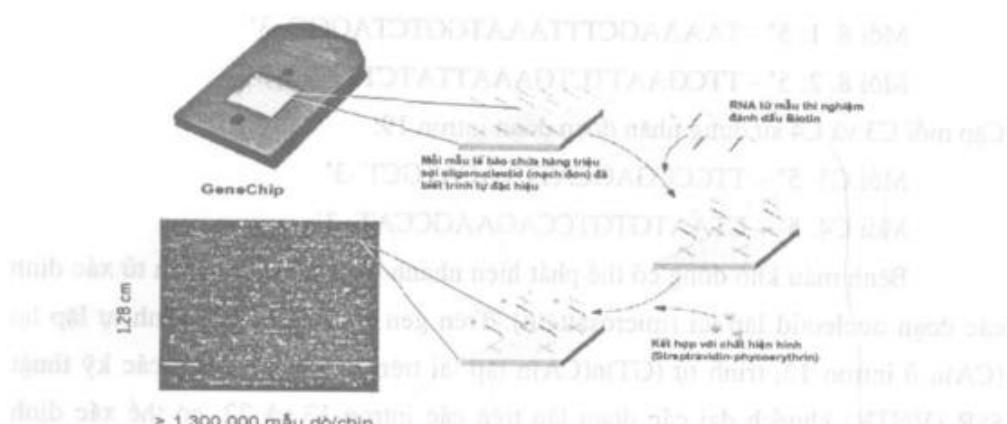
Môи C4: 5' – CTAATGTGTCCAGAACCAT -3'

Bệnh máu khó đông có thể phát hiện nhanh bằng kỹ thuật phân tử xác định các đoạn nucleotid lặp lại (microsatellite). Trên gen F. VIII có các trình tự lặp lại (CA)<sub>n</sub> ở intron 13, trình tự (GT)<sub>n</sub>(CA)<sub>n</sub> lặp lại trên intron 22. Bằng các kỹ thuật SSR (VNTR) khuếch đại các đoạn lặp trên các intron 13 và 22, có thể xác định nhanh người bị bệnh máu khó đông. Ngoài ra, các kỹ thuật Southern Blot, lai huỳnh quang tại chỗ (FISH)... cũng được sử dụng để phát hiện nhanh người mang gen bệnh hoặc người bị bệnh hemophilia.

Năm 1993, Lu và cộng sự điều trị có hiệu quả cho 2 trường hợp bệnh hemophilia bằng sử dụng vectơ virus để chuyển gen F. IX. Thí nghiệm chuyển Adenoviral vectơ mang gen F. IX của người cũng được đưa vào chuột nhắt, làm tăng hàm lượng protein F. IX người trong máu chuột đạt 400 ng / ml (Smith và cộng sự, 1993). Triển vọng chữa bệnh máu khó đông ở người bằng kỹ thuật gen, có thể được thực hiện trong những năm tới.

#### *4. Sử dụng kỹ thuật microarray trong chẩn đoán bệnh*

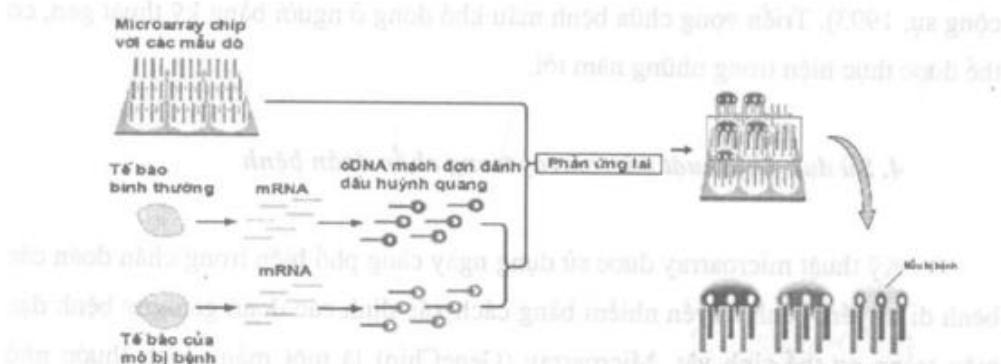
Kỹ thuật microarray được sử dụng ngày càng phổ biến trong chẩn đoán các bệnh di truyền, bệnh truyền nhiễm bằng cách xác định các đoạn gen gây bệnh đặc hiệu trong cơ thể sinh vật. Microarray (GeneChip) là một mảng kích thước nhỏ (thủy tinh hoặc nhựa), chứa một lượng rất lớn các mẫu dò (từ 300 000 đến hàng triệu), cho phép chẩn đoán nhanh và chính xác các bệnh liên quan đến vật chất di truyền (DNA, RNA) trong cơ thể. Mẫu dò là các đoạn DNA nhỏ đã biết triết tự, được đánh dấu huỳnh quang và gắn vào những vị trí xác định trên các mảnh thủy tinh hoặc nhựa, tạo nên các GeneChip.



Hình 6.25. Sơ đồ kỹ thuật microarray trong chẩn đoán

Chẩn đoán bệnh bằng kỹ thuật microarray gồm các bước cơ bản sau:

Tách chiết DNA, nhân gen qua PCR (hoặc tách chiết mRNA, phiên mã ngược và nhân gen tạo cDNA). Biến tính thu các mạch đơn DNA (hoặc cDNA) và đánh dấu huỳnh quang, thực hiện lai phân tử với mẫu dò đặc hiệu của GeneChip.



Hình 6.26. Kỹ thuật microarray trong chẩn đoán ung thư

Phân tích kết quả bằng máy scanner chuyên dụng, nếu các mẫu nghiên cứu mang gen bệnh đặc hiệu, gen đột biến... kết quả hiển thị trên máy tính cho biết các gen bệnh, các gen đột biến và mức độ bệnh tật.

Kỹ thuật microarray có thể sử dụng để chẩn đoán cùng một lúc nhiều loại bệnh truyền nhiễm, bệnh do rối loạn di truyền, bệnh ung thư... với kết quả nhanh, độ chính xác cao.

### *5. Sử dụng kỹ thuật gen trong chẩn đoán bệnh nhiễm trùng ở vật nuôi*

Bệnh nhiễm trùng do virus, vi khuẩn và ký sinh trùng không chỉ gây ảnh hưởng đến sức khỏe con người, còn gây thiệt hại lớn đến hiệu quả sản xuất nông nghiệp. Trong ngành chăn nuôi, đặc biệt ở ngành nuôi trồng thuỷ sản các bệnh do nhiễm virus vi khuẩn làm giảm sản lượng, chất lượng sản phẩm gây thiệt hại lớn cho nền kinh tế. Chẩn đoán sớm và chính xác các bệnh nhiễm trùng, giúp con người tìm được các biện pháp điều trị bệnh, mang lại lợi ích kinh tế lớn trong sản xuất các sản phẩm xuất khẩu và sản xuất nông nghiệp. Kỹ thuật PCR là một trong những kỹ thuật phân tử đang được sử dụng rộng rãi, để chẩn đoán sớm, chẩn đoán nhanh tình trạng nhiễm bệnh do virus, vi khuẩn ở tôm cá trong nuôi trồng thuỷ sản.

Trong nuôi trồng thuỷ sản, nghề nuôi tôm sú có tốc độ phát triển mạnh, chiếm 70 - 90 % tổng giá trị sản phẩm nghề nuôi tôm. Một số vùng nuôi tôm ở qui mô công nghiệp, trên diện rộng ở nhiều nước Đông Nam Á bị thiệt hại rất lớn, do tôm bị nhiễm virus dẫn đến hiện tượng chết hàng loạt (dịch tôm). Một số bệnh thường gặp ở tôm sú như bệnh đốm trắng WSSV (*White spot syndrome virus*), bệnh tôm còi MBV (*Monodom baculovirus*), bệnh tôm bị đầu vàng (YHV).... làm cho tôm chậm lớn, gây chết hàng loạt, làm ô nhiễm môi trường và những tổn thất kinh tế lớn, ảnh hưởng nghiêm trọng đến sự phát triển của nghề nuôi tôm.

Hiện nay có nhiều phương pháp chẩn đoán bệnh do nhiễm virus của tôm sú bằng kỹ thuật phân tử như: phương pháp PCR, phương pháp miễn dịch... Nhiều bộ Kit chẩn đoán nhanh các bệnh nhiễm virus ở tôm sú thương mại hoá được sử dụng rộng rãi. Ví dụ, Kit chẩn đoán tôm sú bị bệnh đốm trắng do nhiễm virus: WSSV Detection Kit (Malaysia), IQ 2000 WSSV (Đài Loan), WSSV MB Tech Kit (ĐH.KHTN. Tp Hồ Chí Minh)....

Kỹ thuật phân tử chẩn đoán bệnh virus ở tôm đã được áp dụng ở nhiều nước Mỹ, Nhật, Úc, Malaysia, Thái Lan, Singapore, Đài Loan... và Việt Nam. Ở nước ta, đã có nhiều công trình nghiên cứu sử dụng kỹ thuật PCR để phát hiện nhanh tôm bị bệnh đốm trắng, bệnh vàng đâu, bệnh còi cọc.....

Sử dụng cặp mồi MBVF: 5' - TCGAATCGCGTTGTACT -3' và MBVR: 5' - CGCTAATGGGGCACAAAGTGTC -3' có thể phát hiện nhanh tôm sú nhiễm *Monodom baculovirus* (Nguyễn Bá Nhựt và cộng sự, 2004).

Kỹ thuật PCR với các cặp mồi đặc hiệu khác nhau, đang được sử dụng rộng rãi ở nhiều nước để phát hiện và chẩn đoán sớm tình trạng nhiễm virus, vi khuẩn ở nhiều loài tôm, cá và các đối tượng khác trong nuôi trồng thuỷ sản, cũng như ở các loại cây trồng trong sản xuất nông nghiệp.

#### **6. Sử dụng kỹ thuật gen trong chẩn đoán bệnh cây trồng và kiểm tra thực phẩm chuyển gen**

Trong trồng trọt, nhiều loại bệnh virus vi khuẩn đã gây thiệt hại rất lớn cho sản xuất nông nghiệp. Ví dụ, bệnh virus Y khoai tây làm giảm 70% - 80% năng suất khoai tây, bệnh héo xanh do vi khuẩn *Xanthomonas* gây giảm sản lượng nhiều loại cây rau quả và cây lương thực như sắn, cà chua, lúa...

Kỹ thuật phân tử được ứng dụng trong chẩn đoán xác định bệnh virus, vi khuẩn sớm, giúp các nhà trồng trọt đưa ra biện pháp phòng trừ hữu hiệu, mang lại hiệu quả to lớn trong sản xuất nông nghiệp.

Kỹ thuật miễn dịch liên kết enzym ELISA (Enzyme Linked Immuno - Sorbent Asay) là một kỹ thuật có độ nhạy cao, được sử dụng kiểm tra và phát hiện sớm nhiều loại bệnh virus (arabis mosaic virus, plumpox virus...) ở cây trồng.

Kỹ thuật PCR sử dụng nhiều cặp mồi khác nhau được ứng dụng chẩn đoán sớm và rất nhiều bệnh do virus và bệnh héo rũ ở cà chua, cây bô đậu, bắp cải do vi khuẩn *Xanthomonas* ...

Kỹ thuật PCR được sử dụng để nhận biết các sản phẩm chuyển gen. Ví dụ, các nhà khoa học Đài Loan sử dụng các cặp mồi 3SS-NOS để phát hiện ngô và

dẫn tương chuyển gen chống cỏ dại. Có thể sử dụng các cặp mồi đặc hiệu với các gen Cry IA (b), CDPK-cry, 35S, CMV promoter... để phát hiện gen Delta endotoxin trong ngô, dẫn tương chuyển gen ...

### Tài liệu tham khảo chính

1. Alm, Gunner V. (2003). Role of natural Interferon-alpha producing cells (Plasmacytoid dendritic cells) in Autoimmunity. *Autoimmunity* 36, 463-472.
2. Couric, John M. (1984). Hemophilia has outlasted the czars. *FDA Consumer. Health* 18.
3. Decatris, Marios. (2002). Potential of Interferon-alfa in Solid Tumours. *Biodrugs* 16, 261-268.
4. Đái Duy Ban và cộng sự (2002). Sinh học phân tử của ung thư vòm họng. NXB Khoa học và Kỹ thuật, Hà Nội.
5. Đái Duy Ban và cộng sự (2004). Sinh học phân tử của các bệnh gan và ung thư gan. NXB Khoa học và Kỹ thuật, Hà Nội.
6. Hertzog, Paul J. (2004) Interferon-gamma: an overview of signals, mechanisms and functions. *Journal of Leukocyte Biology* 75, 163-179.
7. Khuất Hữu Thanh, (2003). Cơ sở di truyền phân tử và kỹ thuật gen. NXB Khoa học và Kỹ thuật, Hà Nội.
8. Khuất Hữu Thanh, (2004). Liệu pháp gen - Nguyên lý và ứng dụng. NXB Khoa học và Kỹ thuật, Hà Nội.
9. Kresina T. F. (2001). An Introduction to Molecular Medicine and Gene Therapy. Wiley-Liss, Inc
10. Lê Đình Công và cộng sự (1996). Chẩn đoán phân biệt 4 loài kí sinh trùng sốt rét trên người bằng kỹ thuật phản ứng chuỗi Polymerase lồng (Nested PCR). Thông tin phòng chống bệnh sốt rét và các bệnh kí sinh trùng. Số 4/1996  
113.190.1.105 downloaded 60930.pdf at Sun Jul 01 23:28:18 ICT 2012

11. Lê Trần Bình và cộng sự (2003). Áp dụng các kỹ thuật phân tử trong nghiên cứu tài nguyên sinh vật Việt Nam. NXB Khoa học và Kỹ thuật, Hà Nội.
12. Lê Đức Đào và cộng sự (2003). Xác định cơ cấu kí sinh trùng sốt rét trong máu và phân biệt tái phát, tái nhiễm bằng kỹ thuật phản ứng chuỗi Polymerase lồng tại tỉnh Gia Lai. Tạp chí phòng chống bệnh sốt rét và các bệnh ký sinh trùng. Số 1/năm 2003.
13. Matthews, James S. (2004). Peginterferon Alfa-2a: A review of approval and investigational uses. Clinical Therapeutics 26, 991-998.
14. Nguyễn Hạnh Phúc (2003). Kỹ thuật PCR-RFLP trong xác định đột biến DNA bệnh Hemophilia A. Báo cáo khoa học Hội nghị Công nghệ sinh học toàn quốc. Hà Nội, 2003, trang 1069-1072.
15. Trần Văn Sáng (2000). Sinh học phân tử và miễn dịch học trong bệnh lý hô hấp NXB Y Học, Hà Nội.
16. Tạp chí Công nghệ Sinh học (2004). Tập 2 (1), trang 33 - 41.
17. Vrolijk, J. M. (2004) The treatment of hepatitis C: history, presence and future. Journal of Medicine 62, 76-82.
19. <http://web4.infotrac.galegroup.com/itw/infomark/>.
20. <http://www3.ncbi.nlm.nih.gov/lhtbin-post/Omim/dispmim?306700>
21. <http://www.expasy.ch/cgi-bin/niceprot.pl?P00451>

## Chương 7

# KỸ THUẬT GEN ỨNG DỤNG TRONG THỰC TIỄN SẢN XUẤT

### I. Kỹ thuật gen và các ứng dụng trong sản xuất nông nghiệp

Sinh vật biến đổi gen (GMO - Genetically Modified Organism) là một vấn đề nhạy cảm, gây nên nhiều tranh luận về các mặt tích cực và tiêu cực trong một thời gian lâu dài. Chủ động tạo nên các giống cây trồng và vật nuôi biến đổi gen theo ý muốn là một trong những thành tựu nổi bật của con người trong lĩnh vực công nghệ sinh học phân tử. Cây trồng và vật nuôi biến đổi gen góp phần tăng năng suất và chất lượng, tăng hiệu quả sản xuất nông nghiệp, bảo đảm an ninh lương thực góp phần giải quyết hiện trạng đói nghèo của nhiều nước chậm phát triển.

#### 1. Cây trồng biến đổi gen

##### 1. 1. Thực trạng cây trồng biến đổi gen

Sinh vật biến đổi gen bao gồm thực vật, động vật và vi sinh vật. Trong đó, cây trồng biến đổi gen có qui mô và tốc độ phát triển mạnh nhất. Năm 1980, những thử nghiệm chuyển gen của vi khuẩn vào cây trồng nhờ *Agrobacterium tumefaciens* được thực hiện.

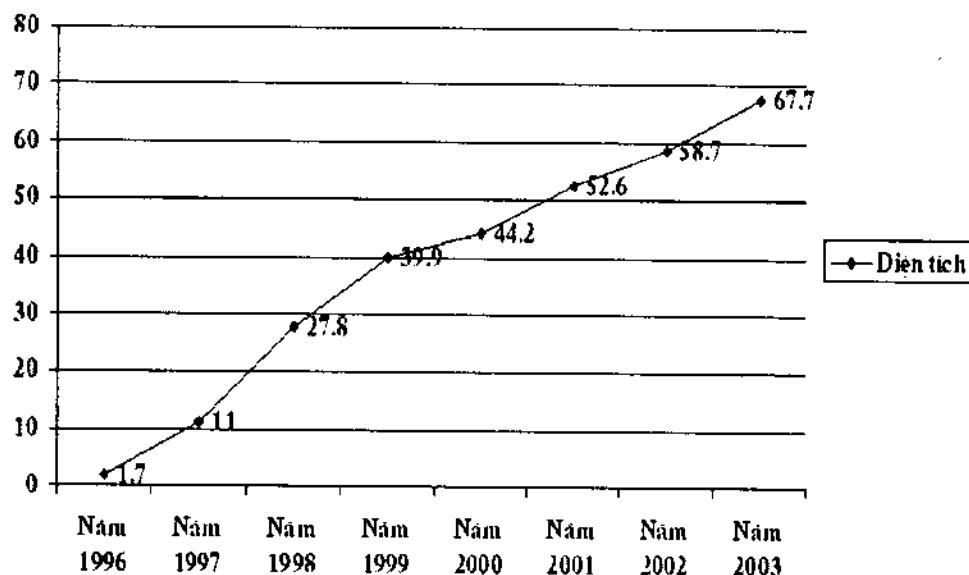
Năm 1986, những cây chuyển gen đầu tiên được trồng thử nghiệm trên đồng ruộng. Năm 1990, thực phẩm chuyển gen đầu tiên được chấp nhận ở Mỹ. Năm 1994,

sản phẩm cây trồng chuyển gen đầu tiên được thương mại hóa là giống cà chua FLAVR SAVR mang gen chín chậm.

Từ 1995, nhiều loại cây trồng biến đổi gen - GMC (Genetically Modified Crop) được đưa ra đồng ruộng, diện tích trồng cây chuyển gen cũng như số loại cây chuyển gen phát triển không ngừng. Trong thời gian 8 năm (1996 - 2003), diện tích trồng cây chuyển gen tăng 40 lần, từ 1,7 triệu ha (1996) tăng lên 67,7 triệu ha (2003). Số quốc gia trồng cây chuyển gen tăng gấp 3 lần, từ 6 nước (1996) đến 2003 có 18 quốc gia trồng cây biến đổi gen.

Năm 2003, diện tích trồng cây chuyển gen vẫn chủ yếu ở các nước phát triển, chiếm 70% tổng diện tích cây chuyển gen toàn cầu với diện tích 42,7 triệu ha. Hiện nay, Mỹ có diện tích trồng cây chuyển gen lớn nhất, chiếm 63 % diện tích cây chuyển gen toàn cầu (42,8 triệu ha), Argentina 13,9 triệu ha (21%), Canada 4,4 triệu ha (6,5), Brazil 3 triệu ha (4%), Trung Quốc 2,8 triệu ha (3,85 %) và các nước Nam Phi có diện tích trồng cây chuyển gen là 0,4 triệu ha chiếm 0,1 % diện tích trồng cây chuyển gen toàn cầu.

**Điện tích cây chuyển gen trên toàn cầu từ 1996 - 2003 (triệu ha)**



Tính đến 2003, đã có hàng chục giống cây chuyển gen được đưa vào sản xuất chủ yếu gồm: đậu tương, ngô, cải dầu (canola), bông, cà chua, du đủ, thuốc lá, cẩm chướng, cúc.... Trong đó, đậu tương chuyển gen có diện tích trồng lớn nhất (41,1 triệu ha chiếm 61%), chủ yếu là đậu tương kháng thuốc trừ cỏ (chịu glufosinate) và đậu tương có hàm lượng acid oleic cao. Các giống ngô chuyển gen kháng sâu, chịu glufosinate... có diện tích trồng lớn thứ 2 với 15,5 triệu ha chiếm 23%, tiếp đó là bông chuyển gen (diện tích đất trồng chiếm 11%)...

Năm 2004, diện tích cây chuyển gen tăng 20% so với năm 2003, đạt 81 triệu ha. Có 8,25 triệu nông dân ở 17 nước trồng cây chuyển gen, trong đó 90% là nông dân ở các nước đang phát triển.

Doanh thu từ cây chuyển gen có mức tăng trưởng rất nhanh, năm 1995 tổng giá trị thị trường toàn cầu của cây chuyển gen đạt 75 triệu USD. Năm 2001 đạt 3,8 tỉ USD, năm 2003 đạt 4,5 tỉ USD - 4,75 USD. Ước tính năm 2005 đạt doanh thu 6,0 tỉ USD, năm 2010 đạt 20,0 tỉ USD. Trong những năm tới, năng suất cây trồng toàn cầu có thể tăng từ 10% đến 25% nhờ phát triển cây chuyển gen.

### 1. 2. Ví dụ ứng dụng kỹ thuật gen tạo cây trồng biến đổi gen

Cây trồng chuyển gen không chỉ được quan tâm đến các gen liên quan với các tính trạng năng suất, tính chống chịu sâu bệnh, chịu các điều kiện môi trường bất lợi (chịu mặn, chịu hạn, chịu lạnh...). Hiện nay, chất lượng sản phẩm cây chuyển gen là vấn đề được chú ý hàng đầu. Các giống cây chuyển gen có chất lượng tốt như cà chua chín chậm có hàm lượng β-caroten cao, đậu tương với hàm lượng acid oleic cao.

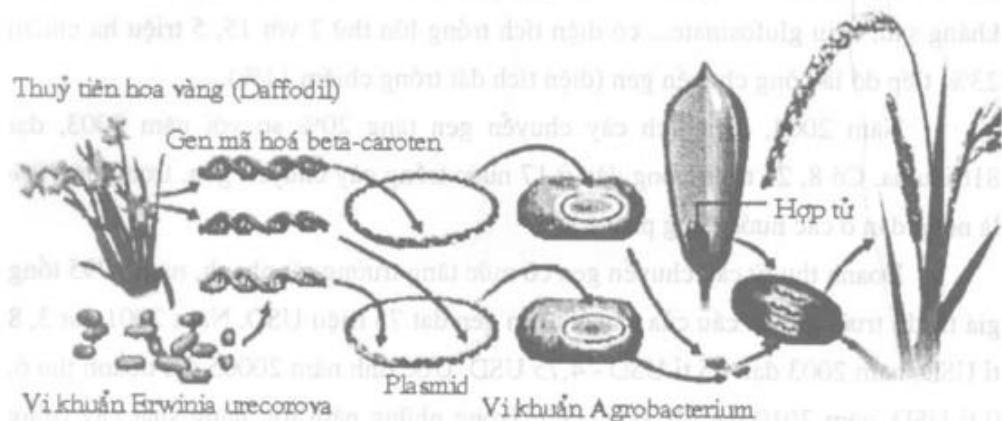
Giống lúa vàng - golden rice là giống lúa chuyển gen có hàm lượng tiền vitamin A (β-caroten), góp phần giải quyết tình trạng thiếu vitamin A của trẻ em trên toàn cầu. Thành tựu tạo giống lúa có sản lượng và hàm lượng β - caroten trong hạt gạo cao, là một bước đột phá trong công nghệ sinh học phân tử nhằm tạo được các giống vật nuôi cây trồng theo nhu cầu con người.

Công nghệ tạo giống lúa vàng - golden rice gồm các bước cơ bản giống như mọi qui trình tạo các cây chuyển gen khác:

Bước 1: Tách chiết các gen mã hoá sự tạo thành  $\beta$ -caroten từ cây Thuỷ tiên hoa vàng (daffodil) và từ khuẩn *Erwinia uredorova*.

Bước 2: Tạo vectơ tái tổ hợp với các plasmid thích hợp và biến nạp vào vi khuẩn *Agrobacterium tumefaciens*.

Thuỷ tiên hoa vàng (Daffodil)



Hình 7.1. Sơ đồ công nghệ tạo giống lúa vàng - golden rice

Bước 3: Chuyển vectơ tái tổ hợp mang các gen mã hoá  $\beta$ -caroten vào các tế bào phôi lúa.

Bước 4: Sàng lọc các tế bào phôi được chuyển gen, nuôi cấy mô để tạo dòng lúa có khả năng tạo  $\beta$ -caroten trong hạt gạo. Sau đó thực hiện lai chéo dòng lúa nuôi cấy mô có hàm lượng  $\beta$ -caroten cao được tạo ra với các giống lúa địa phương nhằm tạo ra giống lúa lai có khả năng tạo  $\beta$ -caroten cao. Tiếp tục cho lai giống để đảm bảo tính ổn định di truyền của các giống lúa có hàm lượng  $\beta$ -caroten cao.

Hiện nay, có vài chục loài cây trồng chuyển gen được thương mại hóa (bông, lúa, lúa mì, khoai tây, ngô, đậu tương, cải bắp, cải dầu, dưa hấu, du dù ớt, nho, cà chua, hạt dẻ...) các giống cây chuyển gen có khả năng chống chịu sâu đục thân, sâu cuốn lá và các loại bệnh khác có hiệu quả rất cao, được trồng phổ biến ở nhiều quốc gia. Nhiều loại cây chuyển gen chống chịu sâu bệnh được đưa vào sản xuất, trong đó

các giống cây được chuyển gen mã hoá protein tinh thể độc của vi khuẩn *Bacillus thuringiensis* (Bt) như ngô, bông, bắp cải... ngày càng tăng mạnh diện tích trồng trọt trên phạm vi toàn cầu.

Sử dụng kỹ thuật gen gây biến đổi một phần cấu trúc gen Cry I A phân lập từ vi khuẩn *Bacillus thuringiensis*, Công ty Monsanto tạo được một gen mới, gen này mã hoá độc tính cao có khả năng diệt côn trùng mạnh. Các nhà khoa học Nga đã thay đổi trình tự các gen Cry I A (b), Cry I A (c) tạo được các gen mã hoá protein độc tố mạnh hơn 10 - 100 lần so với gen cry I A(b) và cry I A(c) tách chiết từ vi khuẩn *Bacillus thuringiensis* tự nhiên.

Ở nước ta, các nghiên cứu tạo giống cây chuyển gen đạt được một số kết quả nhất định, gen kháng bệnh nấm (Chitinase), gen kháng bệnh bạc lá Xa21 đã được chuyển vào một số giống lúa C71, DT10 và DT13. Các giống lúa Nàng Hương, Nàng thơm Chợ Đào biến đổi gen đã được đưa vào sản xuất thử nghiệm có kết quả. Các nhà khoa học Việt Nam đã chuyển thành công gen Cry, gen kháng thuốc diệt cỏ vào nhiều loại cây trồng như bông, thuốc lá, đậu xanh, cải xanh, bắp cải, cà tím.... Các nghiên cứu về cây công nghiệp chuyển gen như cây bông, cây cà phê, cây chè... và nhiều giống rau quả đã thu được một số kết quả bước đầu.

## 2. *Động vật biến đổi gen và động vật nhân bản vô tính*

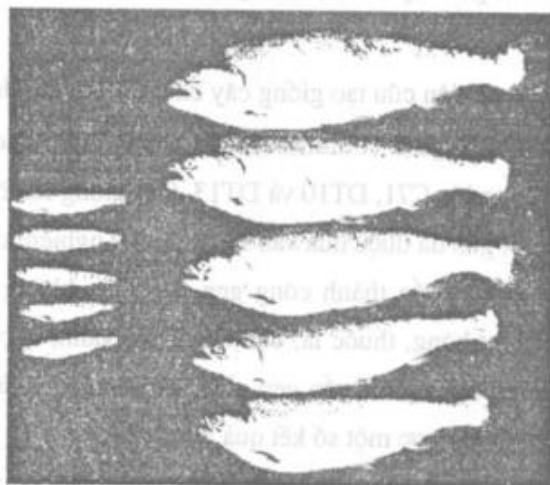
### 2. 1. Tạo giống động vật biến đổi gen mang các đặc tính quý

Các nghiên cứu chuyển gen động vật được thực hiện từ giữa thập kỉ 70, ở nhiều phòng thí nghiệm trên thế giới. Năm 1980, Viện Roslin sử dụng kỹ thuật chuyển gen thành công, tạo cừu biến đổi gen mang gen người đặt tên là Tracy. Trong sữa của cừu Tracy có một loại protein người, loại protein này giúp hạn chế sự phát triển bệnh xơ nang - CF (Cystic fibrosis) ở người.

Năm 1985, Hammer và cộng sự đã chuyển gen hormon sinh trưởng người vào lợn bằng vi tiêm, tạo nên lợn biến đổi gen có tốc độ sinh trưởng nhanh. Hiện nay, các nghiên cứu chuyển gen ở gia súc đã đưa thành công nhiều gen khác nhau

hGH (human growth hormon), bGH (bovine growth hormon), hiGF1 (human insulin - like growth factor 1), PRL (prolactin).... vào bò, dê, cừu lợn... tạo nên các giống vật nuôi biến đổi gen có năng suất cao, sản lượng sữa và phẩm chất thịt tăng đáng kể.

Gen hormon sinh trưởng bò chuyển vào cá hồi (coho salmon), tạo giống cá hồi chuyển gen có tốc độ sinh trưởng cao gấp 11 lần so với cá hồi không chuyển gen (Devlin và cộng sự, 1994).



Hình 7. 2. Cá hồi được chuyển gen hGH và cá hồi bình thường

Bên cạnh các nghiên cứu chuyển gen hormon sinh trưởng vào động vật, tạo động vật chuyển gen có khả năng sinh trưởng cao, người ta đã chuyển một số loại gen mã hoá các đặc điểm quý vào nhiều giống động vật. Các nghiên cứu chuyển gen mã hoá yếu tố VIII của người vào dê, tạo nên giống dê cho sữa có các protein yếu tố VIII (Factor VIII), được ứng dụng trong điều trị bệnh máu khó đông.

Gần đây, tiến sĩ John Clack và cộng sự ở Viện Sinh lý động vật Edinburg đã đưa vào cừu gen mã hoá yếu tố IX của người (một yếu tố cần thiết trong sự đông máu), tạo nên cừu biến đổi gen cho sữa có protein yếu tố IX, có thể tách chiết và sử dụng chữa bệnh máu khó đông hemophilia B ở người. Các nhà khoa học Israel đã

thành công trong việc sản xuất albumin người từ sữa cừu, đây là một phương pháp sản xuất protein máu người hiệu quả cao, đưa lại lợi nhuận rất lớn.

Các nghiên cứu tạo giống lợn biến đổi gen mang gen người, cung cấp nội tạng trong chữa bệnh ở người đã được thử nghiệm thành công ở một số nước. Nội tạng lợn (thận, gan, phổi và tim...) có kích thước tương tự nội tạng người, dễ kiểm giá thành rẻ là đối tượng lí tưởng trong phẫu thuật thay tạng cho người. Sử dụng nội tạng lợn thay cho người gặp rất nhiều khó khăn, do phản ứng miễn dịch loại khỏi. Một khác, trên bề mặt tế bào của các phủ tạng lợn chứa enzym  $\beta$ -galactosidase, không có trong tế bào phủ tạng ở người gây trở ngại cho thay thế phủ tạng.

Năm 1998, các nhà khoa học Anh, Mỹ, Pháp và Nga đã thành công chuyển 2 gen của người vào lợn, tạo giống lợn cho phủ tạng có thể thay thế cho phủ tạng người. Năm 2001 nhóm nghiên cứu Randal Prather tạo đòn lợn chuyển gen có gen  $\beta$ -galactosidase bị bất hoạt, không tạo được enzym  $\beta$ -galactosidase trên bề mặt tế bào của các phủ tạng như ở người. Có thể sử dụng nội tạng lợn thay thế nội tạng cho người không bị thải loại. Thành công này, mở ra ứng dụng rộng rãi sử dụng phủ tạng động vật trong chữa bệnh cho con người, hạn chế mặt tiêu cực của thị trường mua bán nội tạng người.

#### *Ví dụ về công nghệ tạo động vật biến đổi gen*

Động vật biến đổi gen thường được tạo ra bằng các phương pháp chuyển gen hoặc nhân bản vô tính. Phương pháp chuyển gen giúp con người tạo ra các giống động vật có các đặc điểm mong muốn. Hiện nay người ta chủ động chuyển các gen mã hoá hormon sinh trưởng người (hGH) và hormon sinh trưởng bò (bGH)... nhằm tạo các giống động vật chuyển gen có tốc độ sinh trưởng mạnh, năng suất cao.

Qui trình chuyển gen động vật ở giai đoạn phôi là qui trình có hiệu quả cao được sử dụng rộng rãi trong tạo các giống động vật chuyển gen.

Ví dụ công nghệ tạo giống cá mang gen hormon sinh trưởng người, có thể thực hiện gồm một số bước cơ bản sau:

Bước 1: Chọn đối tượng chuyển gen và gen cần chuyển

Đối tượng cần chuyển gen là cá bố mẹ thuần chủng, do các trung tâm giống cá cung cấp.

Gen cần chuyển vào cá là gen mã hoá hormon sinh trưởng người đã được tách dòng trong các vectơ tái tổ hợp, do các hãng sinh phẩm sản xuất. (Ví dụ, có thể mua plasmid pMThGH có kích thước 7,9 kb mang gen hormon sinh trưởng người)

#### Bước 2: Thu nhận phôi cá

Thụ tinh nhân tạo và thu trứng đã thụ tinh, khử màng dính của trứng bằng các dung dịch thích hợp và rửa nhiều lần bằng nước sạch. Loại màng chorion bằng trypsin pha trong Holtfreter nồng độ 0,3%.

#### Bước 3: Chuyển gen

Gen hormon sinh trưởng người (pMThGH) được chuyển vào phôi cá bằng kỹ thuật vi tiêm, nhờ hệ thống thiết bị vi tiêm quan sát dưới kính hiển vi. Có thể sử dụng dung dịch đậm gồm 88 mM NaCl và 10 mM Tris, pH 7,5. Pha loãng plasmid pMThGH nồng độ 100 mg/ml. Trứng sau khi vi tiêm được áp nở trong dung dịch Holtfreter. Phôi được áp trong dung dịch Holtfreter có sục khí và thay nước liên tục để đảm bảo lượng oxi cần thiết đảm bảo tỉ lệ trứng nở và tỉ lệ cá con sống cao.

#### Bước 4: Kiểm tra kết quả chuyển gen

Cá con sau chuyển gen được 3 - 7 ngày tuổi, có thể lấy mẫu vây cá (20 mg) tách chiết DNA và kiểm tra kết quả chuyển gen. Có thể sử dụng cặp mồi đặc hiệu để nhận gen hormon sinh trưởng người trong cá chuyển gen. Trình tự cặp mồi sử dụng cho phản ứng PCR để nhận gen là 5'AAGCGTCACCACGACT3' và 5'AAAAGCCAGGAGCAG3'. Sản phẩm nhận gen hormon sinh trưởng người bằng PCR được kiểm tra bằng phương pháp điện di trên gel agarose 1%, với thang DNA chuẩn để xác định kích thước gen.

Để kiểm tra kết quả chuyển gen chính xác cần thực hiện phương pháp lai phân tử đánh dấu huỳnh quang (FISH), để kiểm tra gen cần chuyển có gắn đúng vị trí cần thiết hay không? kết quả kiểm tra có thể dự báo kết quả chuyển gen.

#### Bước 5: Theo dõi kết quả chuyển gen

Cá con sau khi được chuyển gen, được nuôi trong cùng điều kiện với các con đối chứng. Theo dõi tốc độ sinh trưởng của các chuyển gen để xác định kết quả

chuyển gen. Sau đó có thể thực hiện thụ tinh chéo giống chuyển gen với các không chuyển gen, kết hợp chọn lọc nhiều lần để ổn định di truyền giống cá chuyển gen.

## 2. 2. Nhân bản vô tính động vật

Nhân bản vô tính thực vật được con người ứng dụng rộng rãi trong nuôi cấy mô, tạo các giống cây trồng sạch bệnh, năng suất cao, mang lại nhiều lợi ích cho con người. Nhân bản vô tính động vật được chú ý bắt đầu từ thập kỉ 60. Năm 1962, John Gurdon (Đại học Oxford) đã thành công nhân bản ếch từ những tế bào trứng ếch chuyển nhân khác nhau.

Nhân bản động vật có vú với thành công của Ian Wilmut (1996), đã nhân bản vô tính thành công Dolly, động vật đầu tiên được nhân bản vô tính từ một tế bào trưởng thành. Thành công này tạo nên một chuỗi các nghiên cứu nhân bản vô tính động vật. Một số điểm mốc đánh dấu thành công của nhân bản vô tính động vật, sau thành công nhân bản cừu Dolly:

Năm 1997, trung tâm nghiên cứu động vật có vú (Oregon Regional Primate Research Center) nhân bản thành công hai con khỉ nâu (Mimi và Kiki bằng phương pháp chuyển nhân tế bào phôi).

Năm 1998, bò sữa đầu tiên được nhân bản vô tính trung tâm nghiên cứu được Ishikawa, Nhật Bản.

Năm 2000, con lợn nhân bản vô tính đầu tiên ra đời tại trường đại học Roslin. Năm 2001, bằng nhân bản vô tính tạo nên 5 con lợn mang gen người với các tên Noel, Angel, Star, Joy và Mary nhằm mục đích cung cấp nội tạng thay thế cho con người.

Năm 2002, con mèo có tên là Cc (copy cat) được nhân bản vô tính đầu tiên tại Phòng thí nghiệm Texas. Ngày 26 tháng 12 năm 2002, Công ty Clonaid của giáo phái Rael công bố nhân bản vô tính một đứa trẻ lần đầu tiên với tên là Eva, tuy nhiên không bằng chứng để kiểm tra tính đúng đắn của công bố này.

Năm 2003, con ngựa đầu tiên được nhân bản vô tính.

Đầu tháng 2 năm 2004, Các nhà khoa học Hàn Quốc nhân bản thành công 30 phôi người, các phôi được nuôi giữ nhiều ngày với mục đích nghiên cứu tế bào gốc và sử dụng để điều trị bệnh cho con người.

Tháng 11 năm 2004, các nhà khoa học Trung Quốc đã tiến hành nhân bản thành công gấu trúc, một loài động vật quý hiếm có nguy cơ tuyệt chủng.



**Hình 7.3. Gấu trúc và mèo Cc được nhân bản vô tính**

Hiện nay, nhiều loài động vật đã được nhân bản vô tính thành công. Động vật nhân bản vô tính có vai trò quan trọng trong thực tiễn sản xuất, góp phần tăng nhanh số lượng đàn gia súc, cải tạo di truyền các giống vật nuôi, tạo các giống động vật mang các đặc tính quý phục vụ con người... Thành công nhân bản vô tính động vật, đồng thời mở ra khả năng giúp con người phục hồi các loài động vật quý hiếm đang có nguy cơ tuyệt chủng, có thể phục hồi những loài động vật đã bị tuyệt diệt trên trái đất.

#### **Ví dụ, về công nghệ nhân bản vô tính động vật**

Công nghệ nhân bản vô tính cừu Dolly, là bước nhảy vọt trong công nghệ nhân bản vô tính động vật, mở ra khả năng mới trong cứu hộ các loài động vật có nguy cơ tuyệt chủng hoặc phục hồi một số loài động vật đã tuyệt chủng.

Qui trình nhân bản vô tính cừu Dolly gồm một số bước cơ bản sau:

Bước 1: Chuyển tiếp tế bào

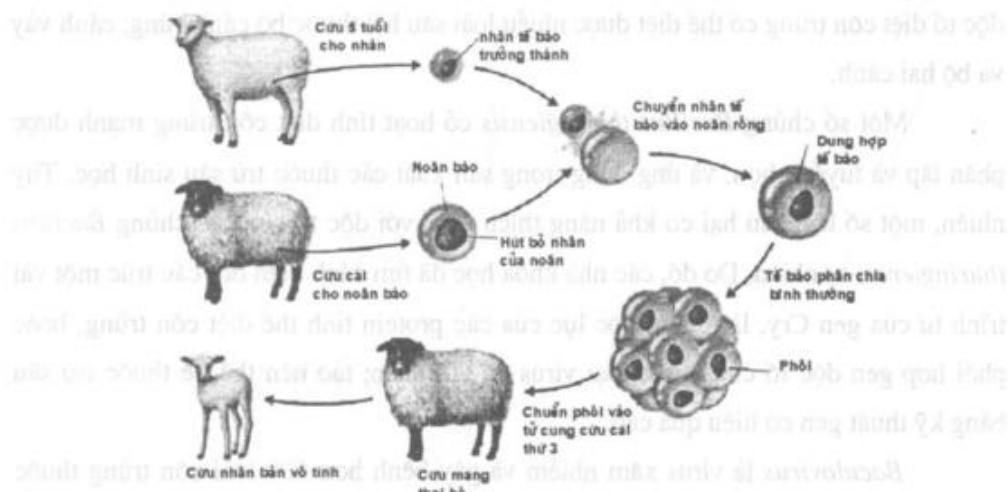
Trong nhân bản cừu Dolly, Wilmut và cộng sự lấy tế bào tuyến vú của một con cừu cái cổ thai (Finn Dorsett), nuôi cấy nhân tạo làm nguồn cung cấp nhân lưỡng bội ( $2n$ ). Các tế bào tuyến vú được nuôi 5 ngày trong điều kiện môi trường tối thiểu (nghèo dinh dưỡng), để các tế bào chuyển sang giai đoạn tinh -  $G_0$  (một giai đoạn nghỉ ngắn trong chu kỳ tế bào). Ở giai đoạn  $G_0$  sự phân chia của tế bào tuyến vú bị kìm hãm, làm đồng bộ hoá nhịp phân chia nhân tế bào soma với nhân của trứng, tạo thuận lợi cho sự phát triển phôi.

#### Bước 2: Loại bỏ nhân của tế bào trứng

Lấy tế bào trứng từ một con cừu cái khác (Black face), sử dụng kỹ thuật vi tiêm để hút bỏ nhân đơn bội ( $n$ ) tạo nên trứng rỗng.

#### Bước 3: Dung hợp nhân tế bào cho với tế bào trứng rỗng

Nhân tế bào tuyến vú và trứng rỗng được đặt vào đĩa Petri, sử dụng xung điện làm dung hợp nhân tế bào tuyến vú với tế bào trứng rỗng tạo thành "hợp tử". Hợp tử bắt đầu phân chia, phôi hình thành và phát triển.



Hình 7.4. Các giai đoạn nhân bản cừu Dolly

#### Bước 4: Cấy hợp tử vào tử cung cừu mẹ mang thai hộ

Hợp tử được nuôi và theo dõi một tuần trong điều kiện thích hợp, cho đến khi hợp tử phát triển thành túi phôi (blastocyst). Lấy túi phôi cấy vào tử cung của một cừu mang thai hộ một con cừu mang thai hộ. Sau 5 tháng, con cừu cái mang thai hộ sinh ra Dolly phát triển bình thường. Dolly có khả năng sinh sản như những cừu cái bình thường khác (Dolly sinh 2 lứa được 4 con). Tuy nhiên Dolly chết sớm, nguyên nhân chết sớm của cừu Dolly đến nay còn chưa hoàn toàn sáng tỏ do sản phẩm nhân bản vô tính, hay đơn thuần do viêm phổi cấp.

### **3. Kỹ thuật gen và thuốc trừ sâu sinh học**

Thuốc trừ sâu sinh học được sản xuất từ các chủng virus, vi khuẩn biến đổi gen ngày càng được ứng dụng có hiệu quả trong bảo vệ thực vật. Vi khuẩn *Bacillus thuringiensis* (Bt) có nhiều chủng khác nhau, mang các gen mã hoá protein tinh thể diệt côn trùng (gen Cry) đặc hiệu khác nhau. Gen Cry ở vi khuẩn *Bacillus thuringiensis* rất đa dạng, gồm các gen từ Cy I đến Cry VI. Các loại protein tinh thể độc tố diệt côn trùng có thể diệt được nhiều loài sâu hại thuộc bộ cánh cứng, cánh vảy và bộ hai cánh.

Một số chủng *Bacillus thuringiensis* có hoạt tính diệt côn trùng mạnh được phân lập và tuyển chọn, và ứng dụng trong sản xuất các thuốc trừ sâu sinh học. Tuy nhiên, một số loại sâu hại có khả năng thích nghi với độc tố của các chủng *Bacillus thuringiensis* tự nhiên. Do đó, các nhà khoa học đã tìm cách biến đổi cấu trúc một vài trình tự của gen Cry, làm tăng độc lực của các protein tinh thể diệt côn trùng, hoặc phối hợp gen độc tố của nhiều loại virus và vi khuẩn, tạo nên thế hệ thuốc trừ sâu bằng kỹ thuật gen có hiệu quả cao.

*Baculovirus* là virus xâm nhiễm và gây bệnh hơn 600 loài côn trùng thuộc các bộ: *Lepidoptera*, *Hymenoptera*, *Diptera* và *Coleoptera*, không gây nhiễm cho người, động vật và thực vật. Bằng cách gắn gen Cry I A(b) và Cry I A(c) vào vector pAcUW2B, sau đó chuyển vào *Baculovirus Autographa California*, các nhà khoa học Nga tạo được *baculovirus* tái tổ hợp. Thuốc trừ sâu sản xuất từ *baculovirus* tái tổ hợp có phổ diệt sâu hại rộng hơn và nhiều lần hiệu lực diệt sâu.

Bằng kỹ thuật thay đổi trình tự cấu trúc các gen Cry I A(a), Cry I A(b) và Cry I A(c), làm tăng độ độc của protein tinh thể, sau đó chuyển vào vi khuẩn *Bacillus thuringiensis*, tạo ra loại thuốc trừ sâu BT có hiệu lực diệt trừ sâu hại cao hơn từ 10 lần đến hàng chục lần, so với thuốc trừ sâu BT sản xuất từ các chủng *Bacillus thuringiensis* hoang dại.

Virus BTV (*Blue Tongue Virus*) mang gen mã hoá một protein độc tố mạnh, khả năng xâm nhiễm côn trùng thấp. Bằng kỹ thuật gen, các nhà khoa học Mỹ đã chuyển gen độc tố mạnh của *Blue Tongue Virus* vào baculovirus, tạo nên loại thuốc trừ sâu virus biến đổi gen có khả năng diệt nhiều loại sâu hại cây trồng với hiệu quả cao.

Hiện nay, nhiều loại thuốc trừ sâu sinh học có nguồn gốc từ virus, vi khuẩn biến đổi gen có hiệu quả diệt sâu cao ngày càng được sử dụng rộng rãi, góp phần nâng cao sản lượng nông nghiệp, đồng thời làm giảm thiểu ô nhiễm môi trường.

## II. An toàn sinh vật biến đổi gen

Vấn đề an toàn của sinh vật biến đổi gen luôn là vấn đề gây tranh luận ở nhiều quốc gia trên thế giới. Cây trồng, động vật biến đổi gen có nhiều ưu điểm rõ rệt như tăng năng suất nông nghiệp, góp phần xoá đói giảm nghèo trên phạm vi toàn cầu, cứu nhiều loài động vật thực vật thoát khỏi nguy cơ tuyệt chủng....

Tuy nhiên, các tranh luận về hiểm họa của sinh vật biến đổi gen vẫn kéo dài, chưa đi đến kết luận rõ ràng. Những nỗi lo sợ sản phẩm biến đổi gen, thực phẩm biến đổi gen (GMF - genetically modified food) gây các dị ứng, mẩn ngứa hoặc ngộ độc cho người sử dụng đã dần dần lắng xuống. Người ta cho rằng việc chuyển gen chống chịu cỏ dại vào cây trồng có thể tạo ra những “siêu cỏ dại”, do thụ phấn chéo giữa cây trồng với cây dại, hoặc chuyển một số gen chống bệnh virus vào động vật, rất có thể gây hiện tượng trao đổi gen với cơ thể người, tạo nên những hiểm họa mới....

Một số các nghiên cứu để kiểm tra protein Cry trong thịt, trứng sữa của các

protein Cry IA (b) trong thịt lợn, thịt bò, thịt gà... Các nghiên cứu đã khẳng định tính an toàn của thực phẩm chuyển gen.

Tổ chức Y tế thế giới (WHO), Liên Hiệp Quốc và nhiều tổ chức kiểm soát lương thực, thực phẩm của Mỹ và thế giới đã khẳng định cây chuyển gen và động vật chuyển gen không gây tác hại đến sức khoẻ con người. Tuy nhiên, một loại cây trồng hoặc động vật chuyển gen trước khi đưa ra sản xuất cần phải được nghiên cứu kĩ và thử nghiệm trong nhiều năm.

Vấn đề nhân bản động vật và nhân bản người cũng gây tranh luận sôi nổi. Một số động vật nhân bản chết khi còn đang được mang thai hoặc chết ngay sau khi ra đời, do những bệnh hô hấp hoặc bệnh về tim mạch. Một số động vật nhân bản thành công như cừu Dolly, cũng có thời gian tồn tại ngắn.... Vấn đề nhân bản người nên hay không nên? Mục đích nhân bản người ? là vấn đề còn được tiếp tục tranh luận. Vào cuối năm 2004, Liên Hợp Quốc tiến hành dự thảo bộ luật cấm nhân bản người, nhằm bảo vệ qua trình tiến hoá tự nhiên của loài người. Dự thảo này đã được 61 quốc gia tán thành, và nhiều quốc gia khác không tán thành. Do còn nhiều tranh luận, nên dự thảo bộ luật cấm nhân bản người đến nay vẫn được tiếp tục thảo luận.

Ở Việt Nam, việc quản lí và sử dụng sản phẩm sinh vật chuyển gen còn một số bất cập, hành lang pháp lí về an toàn sinh học nói chung còn chưa cụ thể. Hiện nay, nước ta là một trong các thành viên chấp hành Công ước Quốc tế về Đa dạng sinh học 1993 (Convention on Biological Diversity), Nghị định thư Cartagena về An toàn sinh học 2000 (Cartagena Protocol on Biosafety). Với Quy chế Quản lí An toàn sinh học đối với sinh vật biến đổi gen và sản phẩm của chúng (Quyết định số /2003/QĐ/TTg) do Chính phủ ban hành năm 2003, và hàng loạt các thông tư liên quan đến an toàn sinh học ở Việt Nam như Luật môi trường, Pháp lệnh kiểm dịch thực vật... đã tạo điều kiện thuận lợi cho sự phát triển mạnh mẽ của công nghệ sinh học cũng như việc ứng dụng các sản phẩm chuyển gen ở nước ta.

**Tài liệu tham khảo chính**

1. Baum J. A, Coyle D. M, Gilbert MP, Jany C. S, Gawron-Burke C. (1990). Novel cloning vectors for *Bacillus thuringiensis*. *Appl Environ Microbiol*, 56(11), 3420- 8.
2. Brown J. R. (2003). Ancient horizontal gene transfer. *Nature Rev. Genet.* 4, 121-132.
3. Chambers J. A, Jelen A, Gilbert M. P, Jany C. S, Johnson T. B, Gawron-Burke C. (1991). Isolation and characterization of a novel insecticidal crystal protein gene from *Bacillus thuringiensis* subsp. *aizawai*. *Bacteriol*, 173(13), 3966 -76.
4. De Cosa B, Moar W, Lee S. B, Miller M., Daniell H. (2001). Over expression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nat Biotechnology*, 19(1), 71- 4.
5. Đái Duy Ban, Lữ Thị Cẩm Vân, (1994). Công nghệ gen và công nghệ sinh học ứng dụng trong nông nghiệp hiện đại. NXB Nông nghiệp.
6. Goodner B. et al. (2001). Genome sequencing of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* 294, 2323-2328.
7. James C. (2001), "Global Review of Commercialized Transgenic Crops: 2000", ISAAA Briefs № 23.
8. James C. (2002), "Global Review of Commercialized Transgenic Crops: 2001", ISAAA Briefs № 24.
9. James C. (1998). Global review of commercialized transgenic crops, ISAAA.
10. Jefferson R. A., Harcourt R. L., Kilian A., Wilson, K. J. & Keese, P. K. (2003). Microbial-glucuronidase genes, gene products and uses there of. US patent 6.391.547.
11. Khuất Hữu Thanh, (2003). Cơ sở di truyền phân tử và kỹ thuật gen. NXB Khoa học và Kỹ thuật, Hà Nội.

12. Martin W. et al. (2003). Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* 99, 12246 - 12251.
13. Nguyễn Đức Thành, (2003). Chuyển gen ở thực vật. NXB Khoa học và Kỹ thuật, Hà Nội.
14. Perry Adkisson và cộng sự, (2004). Genetically Modified Pest-Protected Plants: Science and Regulation. National Academy of Sciences.
15. Phan Cự Nhân, (2001) Di truyền học động vật. NXB Khoa học và Kỹ thuật, Hà Nội.
16. Watson J. D., Baker T. A (2004). Molecular Biology of the Gene. Fifth edition, Pearson Education, Inc., pub. Benjamin Cummings.
17. Van Montagu, Jeff Schell M. (2003): steering *Agrobacterium*-mediated plant gene engineering. *Trends Plant Sci.* 8, 353–354.

# KỸ THUẬT GEN

## Nguyên lý và ứng dụng

Tác giả: PGS. TS. Khuất Hữu Thanh

*Chịu trách nhiệm xuất bản:*

PGS. TS. TÔ ĐĂNG HẢI

*Biên tập và sửa bài:*

ThS. NGUYỄN HUY TIẾN

NGỌC LINH

HOÀNG GIANG

HƯƠNG LAN

*Trình bày bìa:*

**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**

**70 Trần Hưng Đạo - Hà Nội**

---

In 1.000 cuốn, khổ 16 x 24cm, tại: Xưởng in NXB Văn hoá Dân tộc  
Quyết định xuất bản số: 136-2006/CXB/143-06/KHKT ngày 10/5/2006  
In xong và nộp lưu chiểu tháng 5 năm 2006.

206134

TÍkỹ thuật gen



1 007042 801793  
38.000 VND