

# XÂY DỰNG PHẦN MỀM QUẢN LÝ CHỈ THỊ PHÂN TỬ CỦA CÁC GIỐNG VẬT NUÔI VÀ CÂY TRỒNG Ở VIỆT NAM

TS. Phạm Hạ Thủy

Dương Thị Hải Yến


Đại học Nông nghiệp Hà Nội



# NỘI DUNG

- ✓ Mở đầu
- ✓ Sự cần thiết xây dựng CSDL về CTPT các giống vật nuôi & cây trồng của Việt Nam
- ✓ Khái niệm chỉ thị phân tử
- ✓ Nội dung xây dựng
- ✓ Giới thiệu PM quản lý CSDL chỉ thị phân tử của cây lúa ở ĐH Nông Nghiệp HN

# Mở đầu



Trên Thế giới và Việt Nam việc ứng dụng kỹ thuật “chỉ thị phân tử” (molecular marker) vào việc nghiên cứu sinh học, y sinh và đặc biệt trong lĩnh vực nông nghiệp đã đem lại nhiều thành quả to lớn.

Có nhiều loại “chỉ thị phân tử” gắn với một chức năng hoặc sự biến đổi gen có trong một cá thể dẫn tới cá thể đó có những tính trạng nhất định. Việc phát hiện các chỉ thị phân tử này cho ta nhiều thông tin cần thiết đối với một cá thể.

Người ta có thể sử dụng các thí nghiệm để phát hiện các chỉ thị phân tử này. Đối với mỗi chỉ thị phân tử thì có một qui trình thí nghiệm riêng biệt, đôi khi cần chi phí lớn để tìm ra được qui trình thí nghiệm thích hợp.

Việc tập hợp, lưu trữ các thông tin về chỉ thị phân tử cùng với các chỉ số kỹ thuật liên quan trong một cơ sở dữ liệu (CSDL) để thuận tiện cho việc tìm kiếm, sử dụng là hết sức cần thiết.

# Tổng quan tình hình nghiên cứu

## Tình hình nghiên cứu nước ngoài:

- Trong y học: sử dụng để phát hiện và điều trị các bệnh hiểm nghèo như ung thư (Sidransky D, 2002)[3].
- Trong nông nghiệp: lựa chọn các giống lúa kháng bệnh rầy nâu ([Jie Hu](#),2010) [1];
- Có nhiều CSDL chuyên biệt về chỉ thị phân tử được xây dựng tùy theo lĩnh vực áp dụng.
- Viện khoa học quốc gia về nông nghiệp của Hàn Quốc (The National Academy of Agricultural Science (NAAS) đã xây dựng một hệ CSDL các chỉ thị phân tử của cây lúa dựa trên nền web (ChangKug Kim, Ung Han Yoon, GangSeob Lee, 2009)[2]
- Trong năm 2013, Trung tâm Thông tin Công nghệ sinh học nông nghiệp quốc gia (NABIC) của Nhật xây dựng lại một cơ sở dữ liệu chỉ thị phân tử dựa trên web cho nguồn gen hữu ích. [3]
- Ngoài ra còn nhiều các hệ CSDL khác như Kazusa Marker Database chứa CSDL về chỉ thị phân tử của một số giống cây trồng như lạc, khoai tây...hoặc Hê Gramene...

# Tổng quan tình hình nghiên cứu



## Tình hình trong nước:

Trong những năm gần đây, các nhà khoa học chú ý nghiên cứu sử dụng chỉ thị phân tử trong việc chọn các giống lúa thơm (Phan Hữu Tôn, Tống Văn Hải, 2010)[7] ; nghiên cứu tính đa dạng di truyền của đậu tương (Triệu Thị Thịnh, Vũ Thị Thúy Hằng, Vũ Đình Hòa, 2010) [5]; chuẩn đoán bệnh cho tôm he (Nguyễn Chí Thuận, 2005) [6]...

- Phạm Thị Bảy và cộng sự đã phân tích phân tử AND của các dòng lúa chọn lọc đã phát hiện dòng B12 mang đoạn AND dài 1200bp là đoạn AND liên kết chặt chẽ với gen kháng Pi-2(t) và dòng B12 kháng 100% với 4 nòi nấm lây bệnh.

- Lưu Thị Ngọc Huyền, Vũ Đức Quang và cộng sự đã lập bản đồ phân tử cho 3 gen kháng rầy nâu...

Các thông tin về chỉ thị phân tử mới chỉ được lưu trữ tản mát ở các công trình nghiên cứu, chưa có một cơ sở dữ liệu chuyên biệt về chỉ thị phân tử.

# Sự cần thiết phải xây dựng CSDL CTPT của Việt Nam



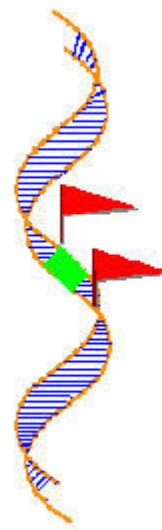
- Nếu có sẵn các CSDL về CTPT thì các nhà nghiên cứu dễ dàng sử dụng các thông tin vào công việc của mình giảm bớt thời gian và chi phí.
- Mặc dù trên thế giới có nhiều CSDL sinh học và chỉ thị phân tử riêng biệt, nhưng việc truy cập ngày càng trở nên khó khăn và phải chi phí.
- Các giống vật nuôi, cây trồng ở Việt Nam có đặc thù riêng do vậy các CTPT cũng có sự khác biệt.
- Cần phải có các CSDL về CTPT riêng của VN để có thể chủ động khai thác, sử dụng...
- Việt Nam có đủ nhân lực và khả năng để xây dựng riêng những CSDL này.



# GIỚI THIỆU VỀ CHỈ THỊ PHÂN TỬ

Chỉ thị phân tử (Molecular Marker) được xác định như chỉ thị di truyền (Genetic Markers) là một gen hoặc đoạn trình tự DNA mà vị trí của chúng đã được biết trên một nhiễm sắc thể mà có thể được sử dụng để xác định đặc tính của các cá thể hay các loài.

## GENETIC MARKERS



"The green section indicates the presence of a desirable gene in an organisms' genetic code that is associated with two genetic markers (red flags)."

Source: <http://usda-ars-beaumont.tamu.edu/dblhelix.jpg>

# The DNA sequences and length of the *Thinopyrum elongatum* 7E chromosome-specific molecular marker M7E\_No.2.

Name	Sequences(5'-3')	Length
M7E_No.2	AGCAAATAAAACGACAGTGCAGCTCGTG- GTTAGTATGAAAATTTACTTTAGTATACTA- CTATCCGCATCTAATGCATGTATGGATGCA- CCAAAATTTGTAATAATAAAGGAGCATTAT- CATATTTGTTTAGCAAGCGAACCGTGGTAC- TTATTGCAGCAGAACACTTCTGAATAAATTC- AATGCGGGAGAGAGGTGTTACCTTCTTAGC- ATTCAGGTAGCTGTCCTTGGGTAGCTCGGT- AAAGGTATTTTTCAAAGGAGTTCTCGACCC- GGTGCTCCATGGTGCAGTATCCAGTGACG- ATGCAATTAGCAGACAGCCCCGGCTGTAA- TT GGTGAAAACC	339 bp

doi:10.1371/journal.pone.0065122.t003

Cấu trúc DNA và chiều dài của chỉ thị phân tử M7E\_No.2 trên nhiễm sắc thể 7E của giống lúa mì dại *Thinopyrum elongatum* – liên quan đến tính kháng rầy và bệnh rỉ sét



# Các kiểu của chỉ thị phân tử

- ✓ RFLP Marker (Restriction fragment length polymorphism markers)
- ✓ SSLP Marker (Simple sequence length polymorphism markers)
- ✓ AFLP Marker (Amplified fragment length polymorphism markers)
- ✓ RAPD Marker (Random amplification of polymorphic DNA markers)
- ✓ VNTR Marker (Variable number tandem repeat markers)
- ✓ SNP Marker (Single nucleotide polymorphism markers)
- ✓ STR Marker (Short tandem repeat markers)

Đối với mỗi kiểu chỉ thị phân tử sẽ có một quy trình thí nghiệm và primer đặc hiệu để phát hiện (có thể bao gồm các bước tách chiết vật liệu di truyền, khuếch đại PCR, phân tích trình tự bằng điện di hoặc phương pháp khác)

# KỸ THUẬT KHUẾCH ĐẠI PCR

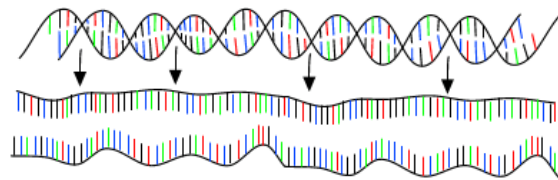


- **PCR** là chữ viết tắt của cụm từ **Polymerase Chain Reaction**, được dịch ở một vài sách là **Phản ứng chuỗi trùng hợp** cũng có sách gọi là "phản ứng khuếch đại gen". PCR là một kỹ thuật phổ biến trong sinh học phân tử nhằm khuếch đại (tạo ra nhiều bản sao) một đoạn DNA mà không cần sử dụng các sinh vật sống như E. coli hay nấm men. PCR được sử dụng trong các nghiên cứu sinh học và y học phục vụ nhiều mục đích khác nhau, như phát hiện các bệnh di truyền, nhận dạng, chẩn đoán những bệnh nhiễm trùng, tách dòng gene, và xác định huyết thống, chọn giống...

# KỸ THUẬT PCR

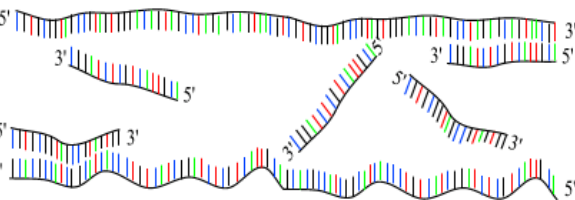
## PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



**Step 1 : denaturation**

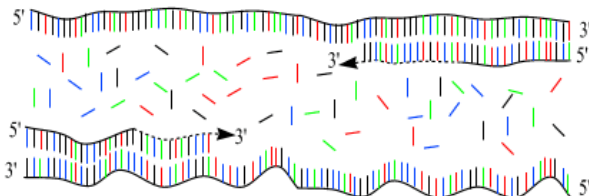
1 minut 94 °C



**Step 2 : annealing**

45 seconds 54 °C

forward and reverse primers !!!

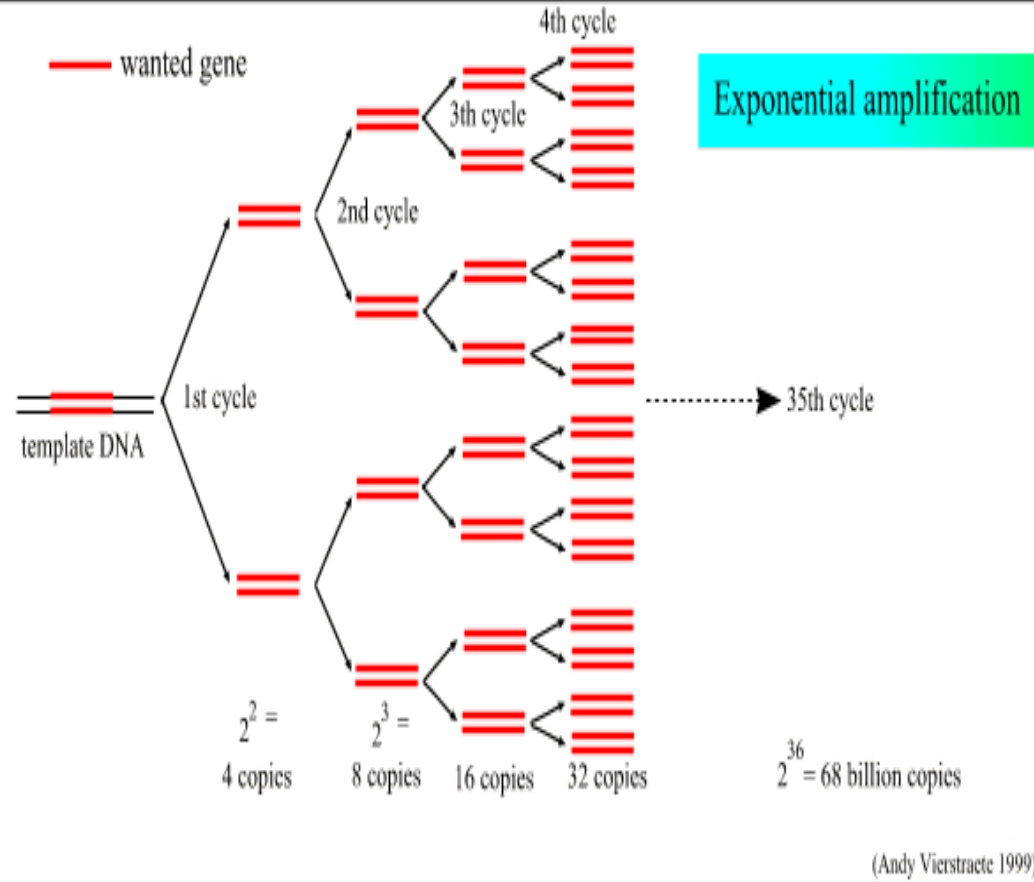


**Step 3 : extension**

2 minutes 72 °C

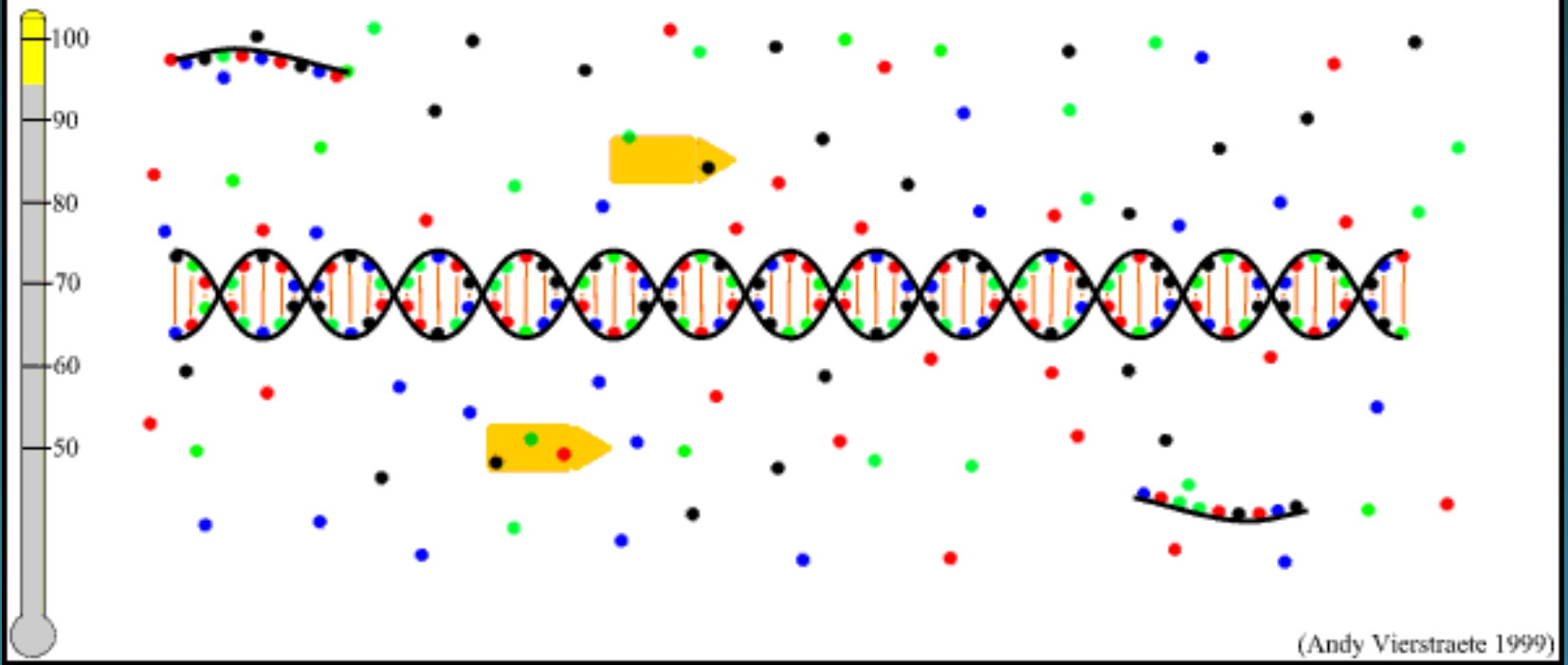
only dNTP's

(Andy Vierstraete 1999)



# KỸ THUẬT PCR

PCR : Denaturation 94°C



<http://users.ugent.be/~avierstr/principles/pcrani.html>



# XÂY DỰNG CƠ SỞ DỮ LIỆU CHỈ THỊ PHẦN TỬ



- MỘT SỐ CSDL CHỈ THỊ PHÂN TỬ CỦA NƯỚC NGOÀI ĐANG ĐƯỢC XÂY DỰNG



# CmMDb: A Versatile Database for *Cucumis melo* Microsatellite Markers and Other Horticulture Crop Research

- Published online 2015 Apr 17. doi: [10.1371/journal.pone.0118630](https://doi.org/10.1371/journal.pone.0118630)
- PMCID: PMC4401682

## Abstract

- The *Cucumis melo* L. microsatellite database (CmMDb: <http://65.181.125.102/cmmdb2/index.html>) encompasses 39,072 SSR markers along with its motif repeat, motif length, motif sequence, marker ID, motif type and chromosomal locations. The database is featured with novel automated primer designing facility to meet the needs of wet lab researchers. CmMDb is a freely available web resource that facilitates the researchers to select the most appropriate markers for marker-assisted selection in melons and to improve breeding strategies

# CerealsDB web-site

- The CerealsDB web-site was created by members of the Functional Genomics Group at the University of Bristol. The site provides a range of facilities for the study of the wheat genome. The site has been designed with breeders in mind, and we hope that it will be easy and straightforward to use. Principally the site is aimed at those who wish to obtain information about SNP markers; e.g., the sequence upon which they are based, obtain primers used for their identification, identify the haplotypes of common UK varieties. These publicly funded resources are available as an on-line, searchable database
- *The last update to the database was February 2015.*

# NABIC marker database: A molecular markers information network of agricultural crops

Chang-Kuo Kim

## • Abstract

- In 2013, National Agricultural Biotechnology Information Center (NABIC) reconstructs a molecular marker database for useful genetic resources. The web-based marker database consists of three major functional categories: map viewer, RSN marker and gene annotation. It provides 7250 marker locations, 3301 RSN marker property, 3280 molecular marker annotation information in agricultural plants. The individual molecular marker provides information such as marker name, expressed sequence tag number, gene definition and general marker information. This updated marker-based database provides useful information through a user-friendly web interface that assisted in tracing any new structures of the chromosomes and gene positional functions using specific molecular markers.

# PlantMarkers—a database of predicted molecular markers from plants

Stephen Rudd<sup>1,\*</sup>, Heiko Schoof<sup>2,3</sup> and Klaus Mayer<sup>3</sup>

- **ABSTRACT**

- Molecular markers are required in a broad spectrum of gene screening approaches, ranging from gene-mapping within traditional 'forward'-genetics approaches through QTL identification studies to genotyping and haplotyping studies. As we enter the post-genomics era, the need for genetic markers does not diminish, even in the species with fully sequenced genomes. PlantMarkers is a genetic marker database that contains a comprehensive pool of predicted molecular markers. We have adopted contemporary techniques to identify putative single nucleotide polymorphism (SNP), simple sequence repeat (SSR) and conserved orthologue set markers. A systematic approach to identify as broad a range of putative markers has been undertaken by screening the available openSputnik

# Xây dựng phần mềm quản lý CSDL về chỉ thị phân tử



## Các thông tin cần lưu giữ trong CSDL

- Chỉ số ID
- Tên của chỉ thị phân tử
- Kiểu chỉ thị phân tử (SNP, RFLP marker...)
- Vị trí trên nhiễm sắc thể
- Trình tự mỗi xuôi (fwd primer seq)
- Trình tự mỗi ngược (rev primer seq)
- Điều kiện cho phản ứng PCR
- Tên của loài (vật nuôi, cây trồng)
- Tính trạng của vật nuôi, cây trồng liên quan
- Tài liệu báo cáo liên quan (tóm tắt hoặc toàn văn)

# Thu thập và tổ chức thông tin



Các thông tin về chỉ thị phân tử được thu thập và chọn lọc từ các nguồn sau:

Từ các báo cáo của các nghiên cứu trong và ngoài nước

Từ các CSDL có sẵn của các tổ chức khoa học có uy tín (ví dụ NCBI, UniProt,...)

Các dữ liệu thông tin thu thập được sẽ được xử lý, rút ra các thông tin cần thiết cho một chỉ thị phân tử được lưu trữ.



# Công cụ xây dựng - Môi trường cài đặt



- ✓ Phần mềm này có thể được xây dựng bằng một ngôn ngữ lập trình web như ASP.NET, PHP...
- ✓ Hệ điều hành: WINDOWS, LINUX...
- ✓ Cài đặt trên một máy chủ được kết nối với mạng nội bộ hoặc internet để cho người dùng (được phép) có thể truy cập vào sử dụng khai thác
- ✓ Các lĩnh vực sử dụng nghiên cứu: chăn nuôi, trồng trọt, nghiên cứu khoa học, giảng dạy và học tập...

# Giới thiệu phần mềm quản lý CTPT của cây lúa

- ✓ Được xây dựng thử nghiệm trong năm 2013.
- ✓ Thông tin được thu thập chủ yếu từ các công trình nghiên cứu của các nhà khoa học trong nước và Đại học Nông nghiệp Hà Nội.

Các chức năng chính của chương trình:

1. Cập nhật lưu trữ các thông tin cơ bản của các chỉ thị CTPT của cây lúa ở Việt Nam
2. Tìm kiếm, tra cứu các CTPT theo kiểu, tên, các công trình nghiên cứu liên quan

Phần mềm đã được cài đặt và chạy thử nghiệm thành công trên máy chủ của Khoa CNTT – ĐH NNHN

# Một số giao diện của phần mềm

## Giao diện chính.

**MOLECULAR MARKERS**  
Quản lý cơ sở dữ liệu về chỉ thị phân tử của cây lúa

[Trang chủ](#) [Tin tức](#) [Tìm kiếm chỉ thị phân tử](#) [Liên hệ](#) [Đăng nhập](#)

**Giới thiệu**

**Proposed Phenotypic Map of the Guppy Y-Chromosome**

Trong di truyền học, một chỉ thị phân tử (Molecular Marker) (được xác định như chỉ thị di truyền (Genetic Markers)) là một gen hoặc đoạn trình tự DNA mà vị trí của chúng đã được biết trên một nhiễm sắc thể mà có thể được sử dụng để xác định các cá thể hay các loài. Nó có thể là một biến thể (có thể phát sinh do đột biến hoặc thay đổi trong các vị trí (locus) của bộ gen) có thể quan sát được. Một chỉ thị di truyền có thể là một chuỗi DNA ngắn, chẳng hạn như là một chuỗi xung quanh một cặp base biến đổi (đa hình đơn nucleotide, SNP), hoặc một chuỗi dài, như các tiểu hành tinh (minisatellites). [Xem tiếp](#) ....

**Liên kết** **Tin tức**

# Giao diện tìm kiếm theo tên CTPT

**Tìm kiếm chi thị phân tử theo tên:**

**Tìm kiếm**

Tên Marker:	AVM22
Kiểu Marker:	AFLP
Vị trí:	2
Trình tự môi xuôi:	GACTGCGTACCAATTCAG
Trình tự môi ngược:	GTAGAGTCCTGAGTAACTT
Điều kiện phản ứng:	94C_4min->(94C_1m)x45->42C_1m->72C_2m->72C_7min->4
Đặc điểm liên kết:	Số lượng rẽ
Tên của loài:	RDB09 và 2021
Kích thước:	200-300
Giống lúa:	Lúa Cạn

**Giới thiệu chung**

**Chi thị phân tử**

**Phần mềm**

**Thư viện**

**Đề tài**

**Tốt nghiệp**

**Nghiên cứu khoa học**

**Báo - Tạp chí**

**Liên kết**



## Tìm kiếm chi thị phân tử theo kiểu

Kiểu chi thị phân tử:

AFLP

**Tìm kiếm**

Tên Marker	Kiểu Marker	Vị trí	Tên của loài	Kích thước	Giống lúa	Chi tiết
AVM22	AFLP	2	RDB09 và 2021	200-300	Lúa Cạn	<a href="#">Chi tiết</a>
AVM38	AFLP	2	RDB09 và 2021	100-200	Lúa Cạn	<a href="#">Chi tiết</a>
AVM62	AFLP	9	RDB09 và 2021	100-200	Lúa Cạn	<a href="#">Chi tiết</a>
<b>AVM8</b>	<b>AFLP</b>	<b>2</b>	<b>RDB09 và 2021</b>	<b>200-300</b>	<b>Lúa Cạn</b>	<b><a href="#">Chi tiết</a></b>
AVM84	AFLP	3	RDB09 và 2021	100-200	Lúa Cạn	<a href="#">Chi tiết</a>
AVM87	AFLP	4	RDB09 và 2021	100-200	Lúa Cạn	<a href="#">Chi tiết</a>

Tên Marker:	AVM8
Kiểu Marker:	AFLP
Vị trí:	2
Trình tự xuôi:	GACTGCGTACCAATTCCT
Trình tự ngược:	GATGAGTCCTGAGTAACTC
Điều kiện phản ứng:	94C_4min->(94C_1m)x45->42C_1m->72C_2m->72C_7m
Đặc điểm liên kết	Tỷ lệ khối lượng rễ/thân
Tên của loài:	RDB09 và 2021
Kích thước:	200-300
Giống lúa:	Lúa Cạn

[Giới thiệu chung](#)

[Chi thị phân tử](#)

[Phân mềm](#)

[Thư viện](#)

[Đề tài](#)

[Tốt nghiệp](#)

[Nghiên cứu khoa học](#)

[Báo - Tạp chí](#)

[Liên kết](#)



# Kết luận

Việc sử dụng kỹ thuật chỉ thị phân tử trong lựa chọn giống vật nuôi và cây trồng cho năng suất cao và chịu chống bệnh là kỹ thuật phổ biến ở nước ngoài, ở Việt Nam mới chỉ bước đầu.

Việc xây dựng các CSDL về CTPP của các giống vật nuôi và cây trồng có ý nghĩa đặc biệt đối với việc phát triển sản xuất nông nghiệp cũng như lĩnh vực nghiên cứu tin sinh học ở VN

Từng bước xây dựng CSDL cho từng giống cây trồng và vật nuôi.

Đòi hỏi sự cộng tác của nhiều nhà nghiên cứu trong và nước ngoài.



# Tài liệu tham khảo

- 
- [1] Jie Hu, Xin Li, Changjun Wu, Changju Yang, Hongxia Hua, Guanjun Gao, Jinghua Xiao, Yuqing He (2010), "Pyramiding and evaluation of the brown planthopper resistance genes Bph14 and Bph15 in hybrid rice", *Molecular Breeding*, [Volume 29, Number 1](#), 61-69, DOI: 10.1007/s11032-010-9526-x
- [2] [Kim ChangKug](#), [Yoon UngHan](#), [Lee GangSeob](#), [Lee HwanKi](#), [Kim YongHwan](#), [Hahn JangHo](#) (2009), "Rice genetic marker database: An identification of single nucleotide polymorphism (SNP) and quantitative trait loci (QTL) marker", *African Journal of Biotechnology* Vol. 8 (13), pp. 2963-2967, 6 July, 2009.
- [3] Kim et al (2013), "NABIC marker database: A molecular markers information network of agricultural crops", *Bioinformatics* 9(17): 887-888 (2013)
- [4] Sidransky D. (2002), "Emerging molecular markers of cancer", [Nat Rev Cancer](#). 2002 Mar;2(3):210-219 .
- [5] Triệu Thị Thịnh, Vũ Thị Thúy Hằng, Vũ Đình Hòa (2010), "Phân tích tính đa dạng di truyền của đậu tương bằng chỉ thị phân tử SSR", *Tạp chí Khoa học và Phát triển* 2010: Tập 8, số 4: 638 - 646.
- [6] Nguyễn Chí Thuận (2005). Nghiên cứu ứng dụng chỉ thị phân tử chẩn đoán bệnh ở tôm he (*Penaeus* spp.) nuôi tại Việt Nam. Kỷ yếu hội thảo toàn quốc về nghiên cứu và ứng dụng khoa học công nghệ nuôi trồng thủy sản, 2005/Số 0, 751 - 759 .
- [7] Phan Hữu Tôn, Tống Văn Hải (2010). Sàng lọc các giống lúa có chứa gen mùi thơm bằng chỉ thị phân tử . *Tạp chí Khoa học và Phát triển* 2010: Tập 8, số 4: 646 - 652

# Thank You !





- Bp base pair
- Mer: đơn vị đo chiều dài
- 10-mer: độ dài 10 base
- Mas: Marker-assisted selection

