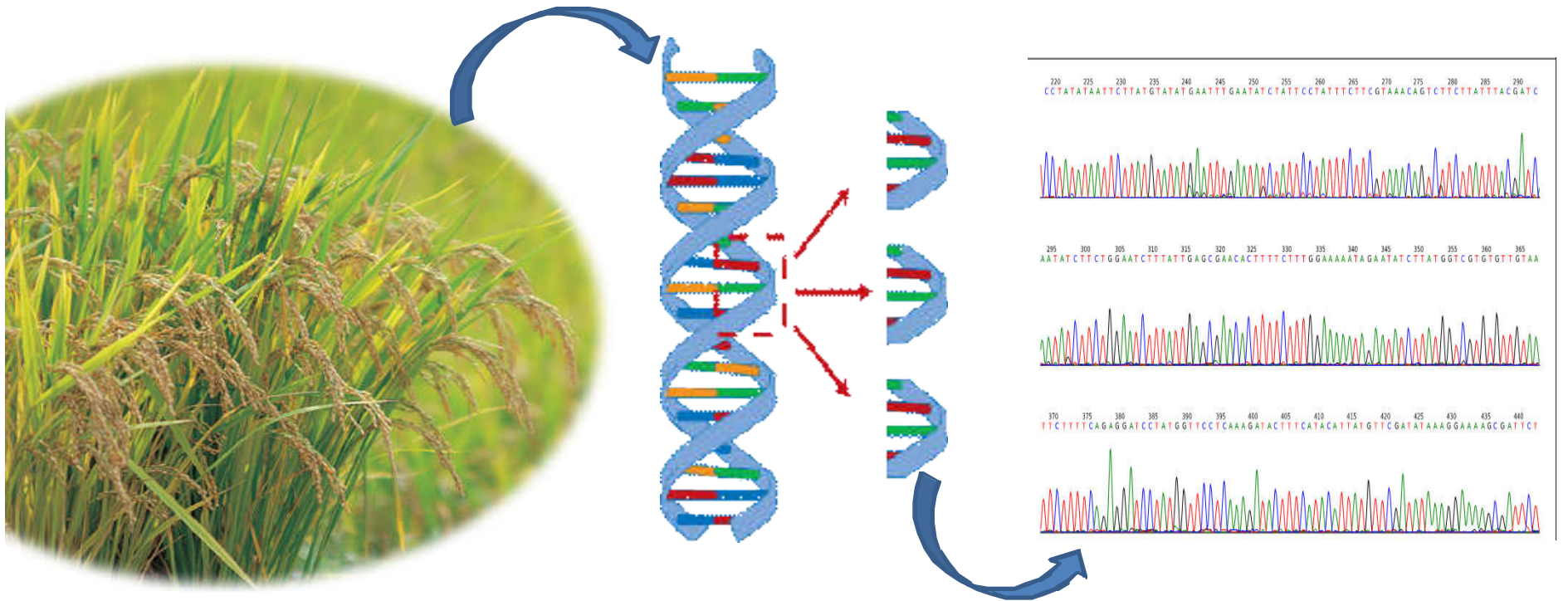




# VIỆN KHOA HỌC NÔNG NGHIỆP VIỆT NAM VIỆN DI TRUYỀN NÔNG NGHIỆP

## GENOME HỌC: HIỆN TRẠNG VÀ TRIỂN VỌNG



*Drought tolerant rice*



*Salt tolerant rice*



*Blast rice*



*Brown planthopper*



*Blight bacterial*



*High quality rice*

# NỘI DUNG BÁO CÁO

---

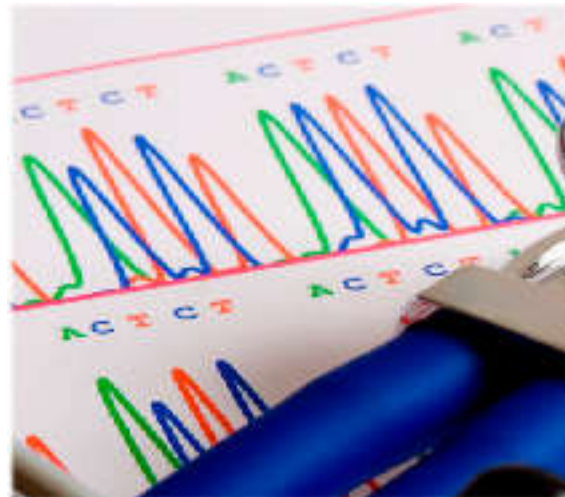
I. CÔNG NGHỆ OMICs

II. TÌNH HÌNH GIẢI MÃ GENOME TRÊN THẾ GIỚI

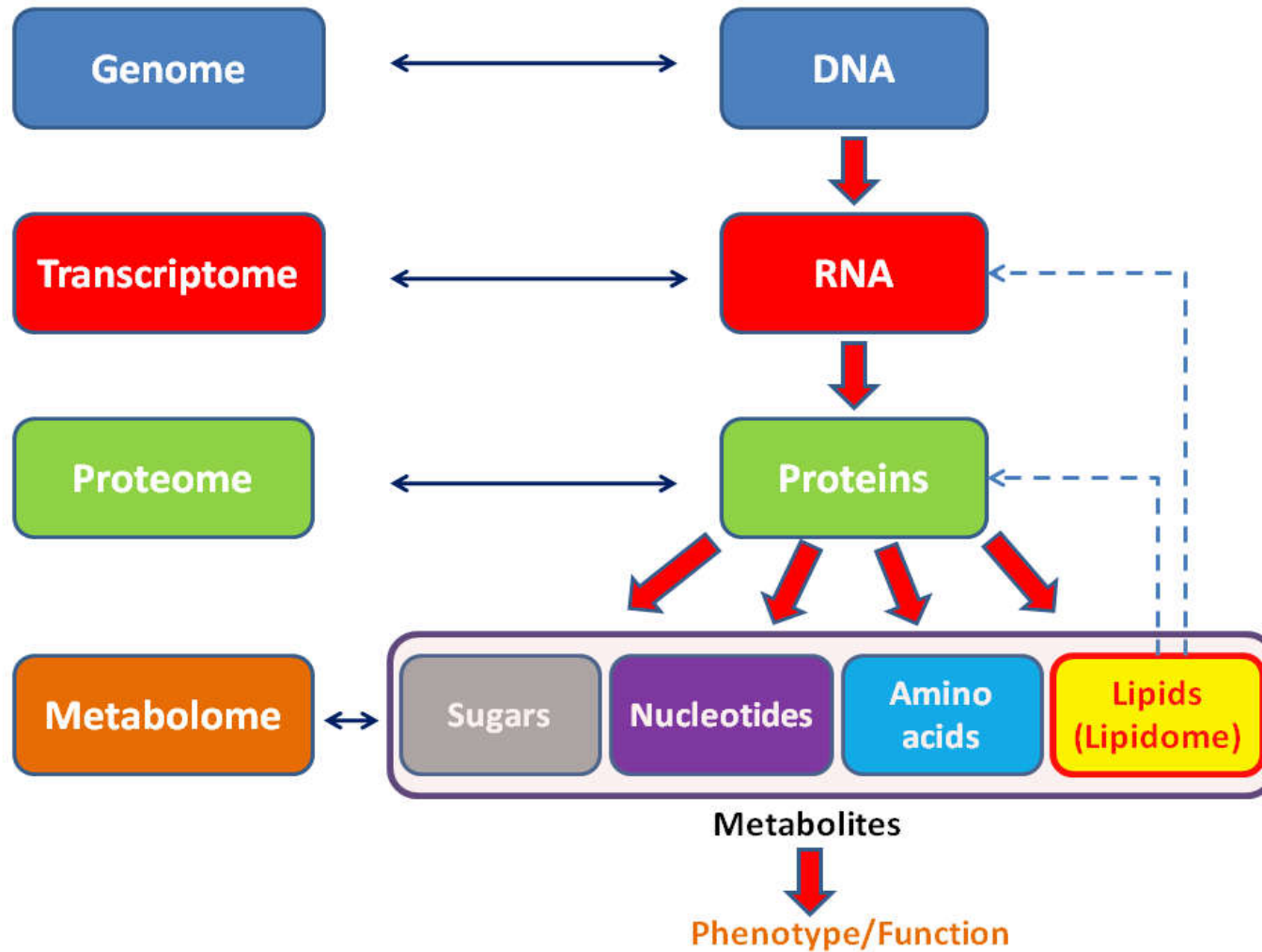
III. CÔNG CỤ GENOME TRONG CHỌN TẠO GIỐNG CÂY TRỒNG

IV. TÌNH HÌNH GIẢI MÃ GENOME TẠI VIỆT NAM

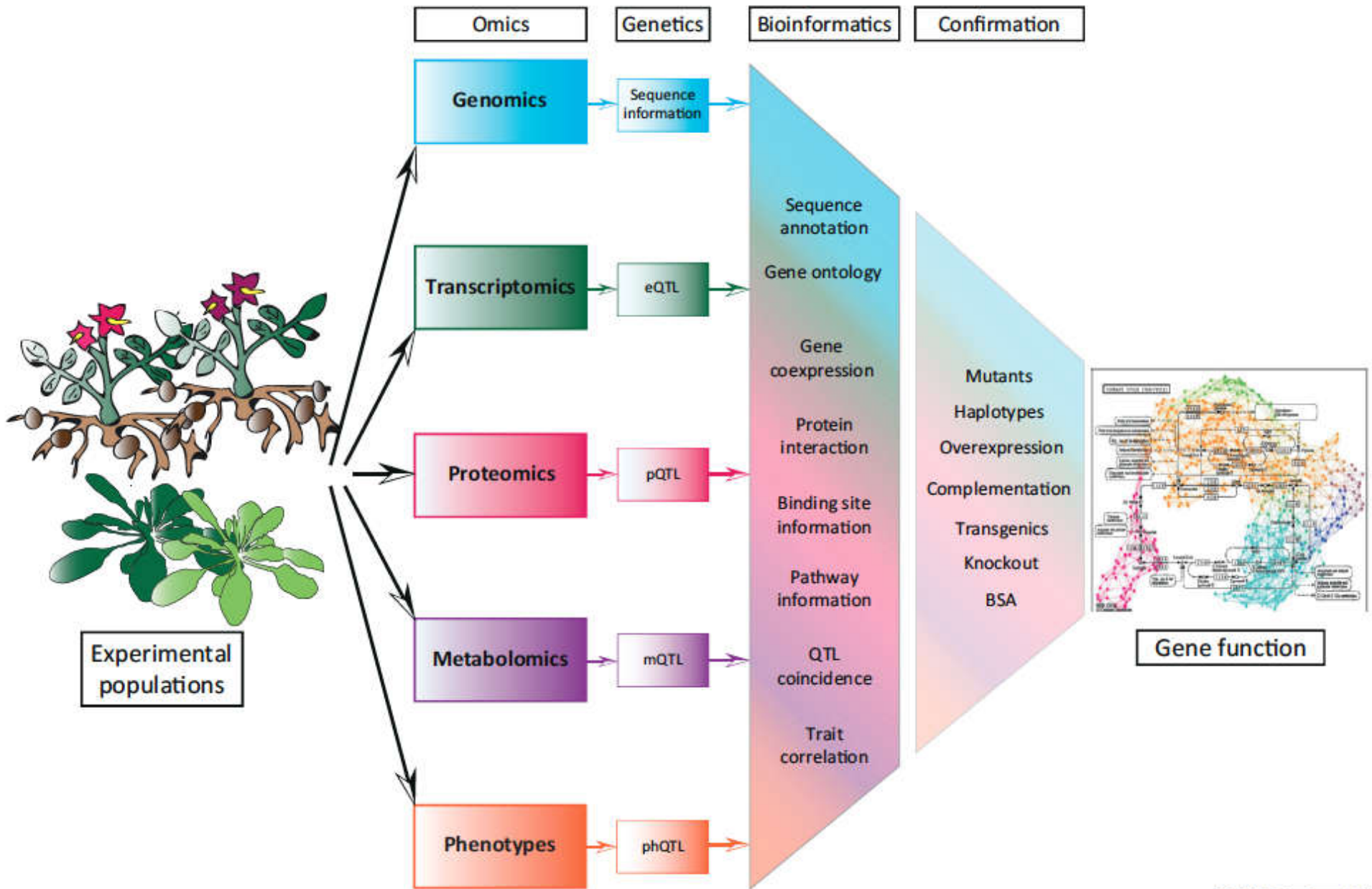
V. KẾT QUẢ GIẢI MÃ GENOME LÚA (AGI-UK), ĐỊNH HƯỚNG NGHIÊN CỨU, KHAI THÁC VÀ ỨNG DỤNG



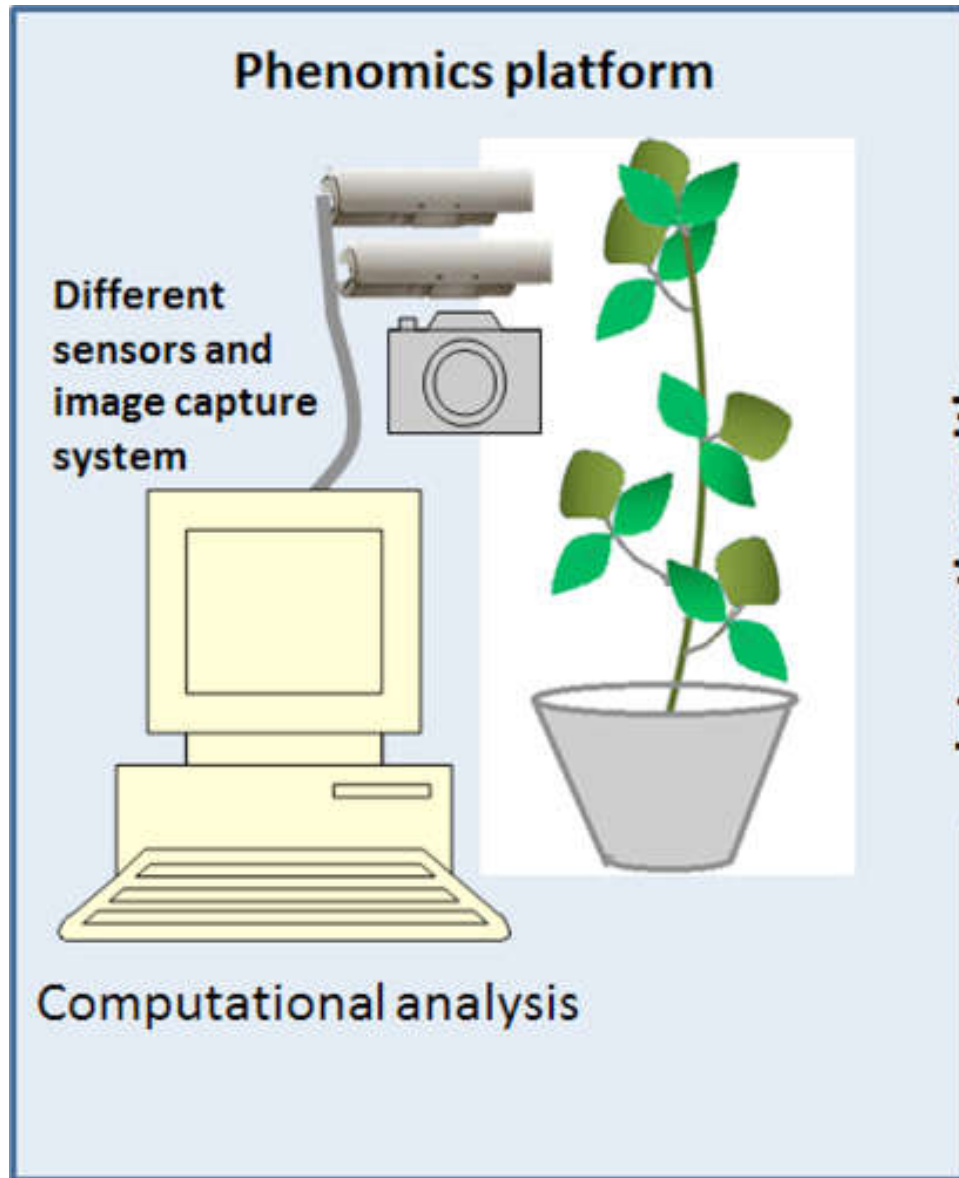
# Omics



# Omics Technologies



# Omics Technologies



Integration with  
Omic approaches

## Genomics

- Gene identification
- Transgene effect
- Genetic control

## Transcriptomics

- Gene regulation
- Expression effect
- Gene regulation

## Proteomics

- Protein identification
- Protein effect
- Molecular mechanism

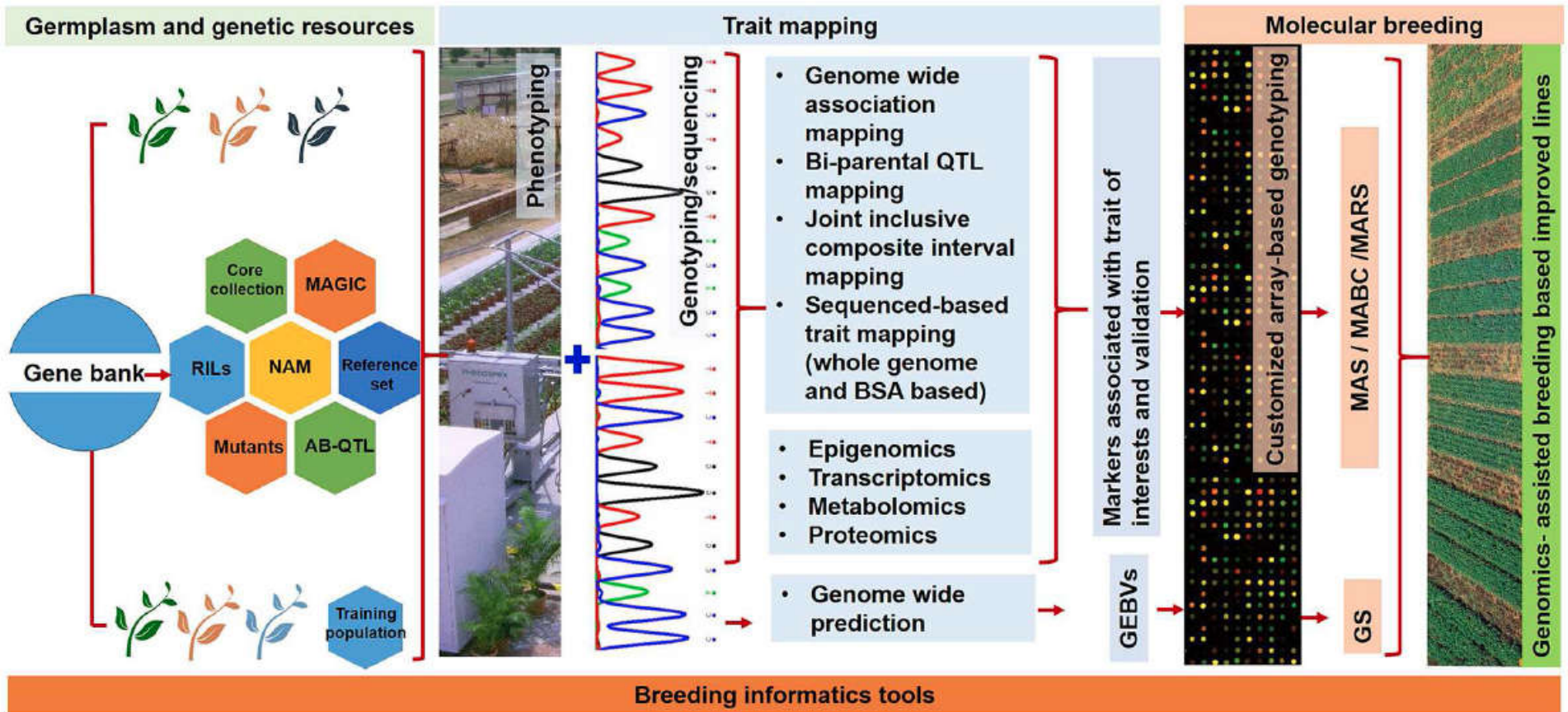
## Metabolomics

- Metabolic pathway
- Metabolite effect
- Regulation

## Inomics

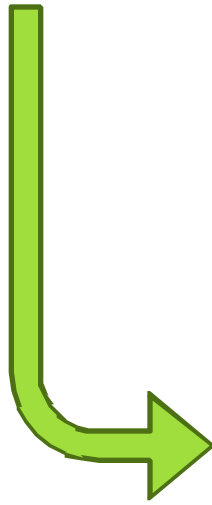
- Elements identification
- Effect of elements
- Plant physiology

# Genomics – assisted breeding



# GWS vs. phenotypic selection

Training Population  
genotyped & phenotyped



parents

	1	2	3	4	5	6	7	8	9	10	11
1		X		X	X						
2			X		X	X					
3	X					X	X				
4		X			X			X			
5			X			X			X		
6				X						X	X

parents

Genotyping & apply  
prediction equation

~ 0.5 year  
low cost

plantation  
phenotypic scoring

≥ 4 years  
high cost

**Selection**

Genomic  
Breeding Value

Phenotypic  
Mean

**TRUE BREEDING  
(or GENETIC) VALUE**

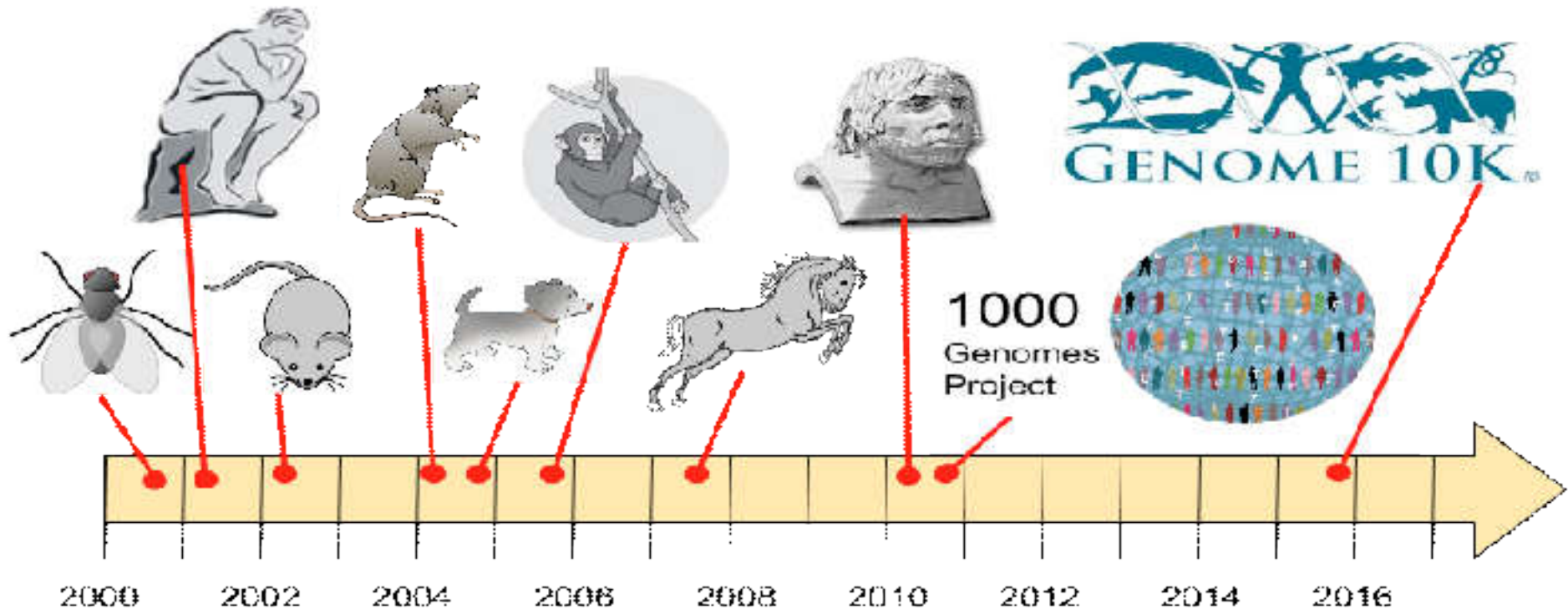
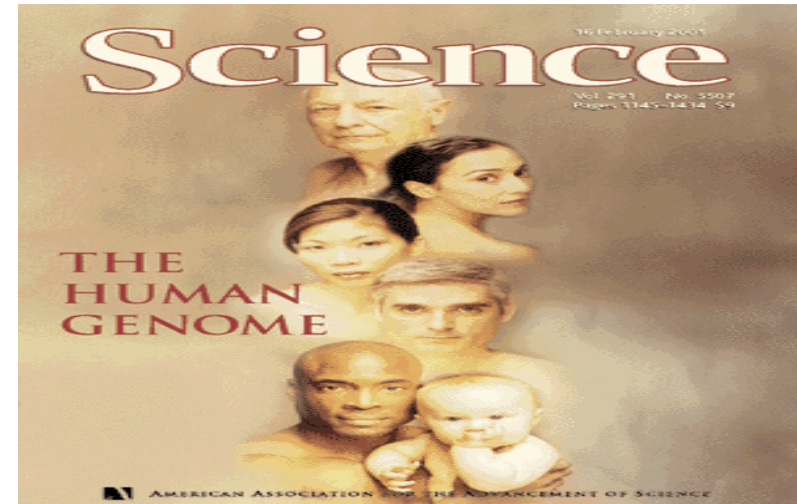
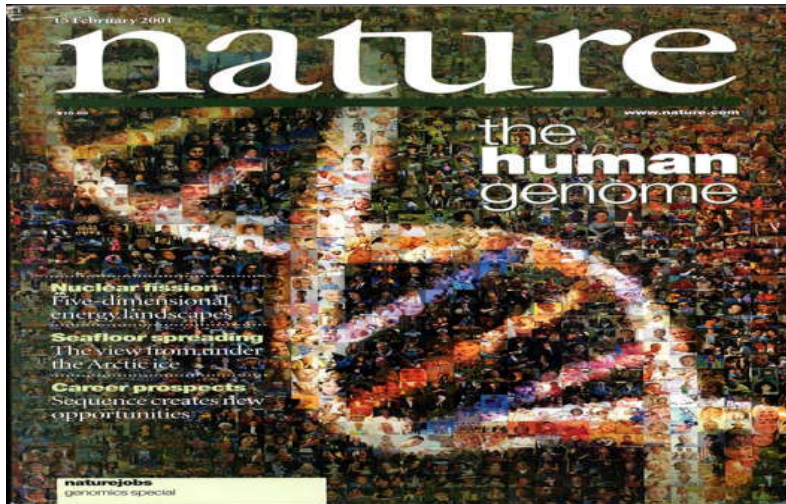


# **I. TÌNH HÌNH GIẢI MÃ GENOME TRÊN THẾ GIỚI**



# Giải mã genome người và động vật

2/2001 – Công bố bản phác thảo đầu tiên về trình tự hệ gen người



# Current research: Comparative genomics

## Some of the animals whose genomes are being or have been sequenced:



Platypus  
*Ornithorhynchus anatinus*



Dog  
*Canis familiaris*  
Photo: Courtesy of The Broad Institute of MIT and Harvard



African elephant  
*Loxodonta africana*



Fruitfly  
*Drosophila melanogaster*



Silkworm  
*Bombyx mori* Daza  
Photo: Alden M. Johnson, California Academy of Sciences



Cow  
*Bos taurus*  
Photo: Courtesy of Terri Hobbs (www.crazyforcows.com)



European rabbit  
*Oryctolagus cuniculus*



Honey Bee  
*Apis mellifera*  
Photo: Courtesy of Scott Bauer, USDA/ARSLaboratories



Armadillo  
*Dasypus novemcinctus*  
Photo: Courtesy of B. Bagatto, Department of Biology, University of Akron



Guinea pig  
*Cavia porcellus*



Roundworm  
*Caenorhabditis elegans*  
Photo: Courtesy of Erik Jorgensen, University of Utah



Zebrafish  
*Danio rerio*



Mouse  
*Mus musculus*  
Photo: Courtesy of Jackson Laboratories



Chicken  
*Gallus gallus*



Opossum  
*Monodelphis domestica*  
Photo: Courtesy of Don Sakaguchi



Cat  
*Felis domesticus*  
Photo: Courtesy of Dr. Kristina Narfstrom, University of Missouri-Columbia



# Algae (Tảo)

Organism strain	Genome size	Organization	Year of completion
<a href="#">Cyanophora paradoxa</a>		<a href="#">Rutgers University</a> <sup>[1]</sup>	2012 <sup>[11]</sup>
<a href="#">Bathycoccus prasinus</a> BBAN7	15 Mb	<a href="#">Joint Genome Institute</a>	2012 <sup>[2]</sup>
<a href="#">Chlamydomonas reinhardtii</a> CC-503 cw92 mt+	111 Mb	University of California at Los Angeles <sup>[3]</sup>	2007
<a href="#">Coccomyxa subellipsoidea</a> sp. C-169		<a href="#">Joint Genome Institute</a>	2007 <sup>[5]</sup>
<a href="#">Dunaliella salina</a> CCAPI9/18		<a href="#">Joint Genome Institute</a>	
<a href="#">Micromonas pusilla</a> CCMP1545		<a href="#">Joint Genome Institute</a>	2007 <sup>[7][8]</sup>
<a href="#">Micromonas pusilla</a> RCC299/NOUM17		<a href="#">Joint Genome Institute</a>	2007 <sup>[8][9]</sup>
<a href="#">Ostreococcus lucimarinus</a> CCE9901	13.2 Mb		2007 <sup>[10]</sup>
<a href="#">Ostreococcus</a> sp. RCC809		<a href="#">Joint Genome Institute</a>	2008 <sup>[12]</sup>
<a href="#">Volvox carteri</a>	~131.2 Mb		2010 <sup>[13]</sup>
<a href="#">Chondrus crispus</a>	105 Mb	Genoscope/Station Biologique de Roscoff	2013 <sup>[14]</sup>
<a href="#">Cyanidioschyzon merolae</a> Strain:10D	16.73 Mb		2004, <sup>[15]</sup> 2007 <sup>[16]</sup>
<a href="#">Galdieria sulphuraria</a>	13.7 Mb		2005 <sup>[17]</sup> 2005 <sup>[18]</sup> 2013 <sup>[19]</sup>
<a href="#">Porphyridium purpureum</a>	19.7 Mb		2013 <sup>[20]</sup>
<a href="#">Pyropia vezoensis</a>	43 Mb		2013 <sup>[21]</sup>
<a href="#">Ectocarpus siliculosus</a>		Station Biologique de Roscoff	2010 <sup>[22]</sup>

# Monocots (thực vật một lá mầm)

## Grasses (Họ Hòa thảo)

Organism strain	Genome size	Organization	Year of completion
<a href="#">Aegilops tauschii</a> (Tausch's goatgrass)	ca 4.36Gb	BGI	2013 <sup>[89]</sup>
<a href="#">Hordeum vulgare</a> (barley)		IBSC	2012 <sup>[91]</sup>
<a href="#">Oryza rufipogon</a> (red rice)	406 Mb	SIBS	2012 <sup>[94]</sup>
<a href="#">Oryza sativa</a> (long grain rice) ssp indica			2002 <sup>[95]</sup>
<a href="#">Oryza sativa</a> (Short grain rice) ssp japonica			2002 <sup>[96]</sup>
<a href="#">Panicum virgatum</a> (switchgrass)			2013? <sup>[97]</sup>
<a href="#">Phyllostachys edulis</a> (moso bamboo)			2013 <sup>[98]</sup>
<a href="#">Sorghum bicolor</a> genotype BTx623	ca 730Mbp		2009 <sup>[99]</sup>
<a href="#">Triticum urartu</a>	ca 4.94Gb	BGI	2013 <sup>[101]</sup>
<a href="#">Zea mays</a> (maize) ssp mays B73	2,300Mbp		2009 <sup>[103]</sup>

# Monocots (thực vật một lá mầm)

## Other non-grasses

Organism strain	Genome size	Organization	Year of completion
<a href="#">Musa acuminata</a> (Banana)	523 Mbp		2012 <sup>[104]</sup>
<a href="#">Musa balbisiana</a> (Wild banana)	438 Mbp		2013 <sup>[105]</sup>
<a href="#">Phoenix dactylifera</a> (Date palm)	658 Mbp		2011 <sup>[106]</sup>
<a href="#">Elaeis guineensis</a> (African oil palm)	~1800 Mbp		2013 <sup>[107]</sup>
<a href="#">Spirodela polyrrhiza</a> (Greater duckweed)	158 Mbp		2014 <sup>[108]</sup>

# Rosids (Thực vật hai lá mầm)

Organism strain	Genome size	Organization	Year of completion
<a href="#">Betula nana</a> (dwarf birch)	450 Mbp	<a href="#">QMUL/SBCS</a>	2013 <sup>[31]</sup>
<a href="#">Arabidopsis thaliana</a> Ecotype:Columbia	135 Mbp		2000 <sup>[34]</sup>
<a href="#">Capsella rubella</a>	130Mbp	JGI	2013? <sup>[36]</sup> 2013 <sup>[37]</sup>
<a href="#">Eutrema salsugineum</a>	240Mbp	JGI	2013 <sup>[38]</sup>
<a href="#">Cannabis sativa</a> (hemp)	ca 820Mbp		2011 <sup>[40]</sup>
<a href="#">Carica papaya</a> (papaya)	372Mbp		2008 <sup>[41]</sup>
<a href="#">Citrullus lanatus</a> (watermelon)	ca 425Mbp	BGI	2012 <sup>[43]</sup>
<a href="#">Cucumis melo</a> (Muskmelon) DHL92	450Mbp		2012 <sup>[44]</sup>
<a href="#">Cucumis sativus</a> (cucumber) 'Chinese long' inbred line 9930	350 Mbp (Kmer depth) 367 Mbp (flow cytometry)		2009 <sup>[45]</sup>
<a href="#">Manihot esculenta</a> (Cassava)	~760Mb	JGI	2012 <sup>[48]</sup>
<a href="#">Ricinus communis</a> (Castor bean)	320Mbp	JCVI	2010 <sup>[49]</sup>
<a href="#">Glycine max</a> (soybean) var. Williams 82	1115Mbp		2010 <sup>[55]</sup>
<a href="#">Phaseolus vulgaris</a> (common bean)	520Mbp	JGI	2013? <sup>[58]</sup>
<a href="#">Linum usitatissimum</a> (flax)	~350Mbp	BGI et al.	2012 <sup>[59]</sup>
<a href="#">Azadirachta indica</a> (neem)	364 Mbp	<a href="#">GANIT Labs</a>	2012 <sup>[64]</sup> and 2011 <sup>[65]</sup>
<a href="#">Fragaria vesca</a> (wild strawberry)	240Mbp		2011 <sup>[67]</sup>
<a href="#">Malus domestica</a> (apple) "Golden Delicious"	~742.3Mbp		2010 <sup>[68]</sup>
<a href="#">Prunus persica</a> (peach)	265Mbp		2013 <sup>[71]</sup>
<a href="#">Populus trichocarpa</a> (poplar)	510 Mbp (cytogenetic) 485 Mbp (coverage)		2006 <sup>[75]</sup>

# Một số cây trồng (Asterids)

Organism strain	Genome size	Organization	Year of completion
<a href="#">Mimulus guttatus</a>	ca 430Mbp	JGI	2013? <sup>[77]</sup>
<a href="#">Solanum lycopersicum</a> (tomato) cv. Heinz 1706	ca 900Mbp	SGN	2011 <sup>[78]</sup> 2012 <sup>[79]</sup>
<a href="#">Solanum tuberosum</a> (potato)	844 Mbp kmer (856 Mbp)	PGSC	2011 <sup>[80]</sup>
<a href="#">Solanum commersonii</a> (commerson's nightshade)	838 Mbp kmer (840 Mbp)	UNINA, UMN, UNIVR, Sequentia Biotech, CGR	2015 <sup>[81]</sup>
<a href="#">Nicotiana benthamiana</a>	ca 3Gbp		2012 <sup>[82]</sup>
<a href="#">Nicotiana glauca</a> (Tobacco plant)	2.636Gbp	Philip Morris International	2013 <sup>[83]</sup>
<a href="#">Nicotiana glauca</a> <i>tomentosiformis</i>	2.682 Gb	Philip Morris International	2013 <sup>[83]</sup>
<a href="#">Capsicum annuum</a> (Pepper) (a) cv. CM334 (b) cv. Zunla-1	~3.48 Gbp		(a) 2014 <sup>[84]</sup> (b) 2014 <sup>[85]</sup>
<a href="#">Capsicum annuum</a> var. <a href="#">glabrusculum</a> (Chiltepin)	~3.48 Gbp		2014 <sup>[85]</sup>
<a href="#">Utricularia gibba</a> (humped bladderwort)	81.87 Mb	LANGEBIO, CINVESTAV	2013 <sup>[87]</sup>

## Những mốc thời gian và tiến bộ trong giải mã genome

### Khung thời gian và chi phí giải trình tự (Sanger):

- **1997** genome nấm men 13Mb ~10 năm; chi phí \$30M
- **1988-2003** genome người 3Gb – 15 năm; \$2.7 billion (1991)

### Giải trình tự hệ gen (20 x coverage) in **9/ 2008** (2 Illumina GAI):

- Genome nấm men 13Mb – 8 genomes per 6 day run; ~\$1,2k/genome
- Genome người ~3Gb – 10-15 tuần; ~\$150k

### Giải trình tự hệ gen (20 x coverage) in **3/ 2009** (3 Illumina systems GAI):

- Genome người ~3Gb: 7 tuần; ~\$75k (rẻ hơn 10,000 lần)

### Giải trình tự hệ gen (20 x coverage) in **3/ 2011** (1 Illumina Hi-seq):

- Genome người ~3Gb: 10 ngày; chi phí ~\$15k (rẻ hơn 50,000 lần, nhanh hơn 500 lần) . **1/ 2012** : \$3-5k, nhanh hơn 1000 lần

### Giải trình tự hệ gen (20 x coverage) in **2015** (1 Illumina Hi-seq 2500):

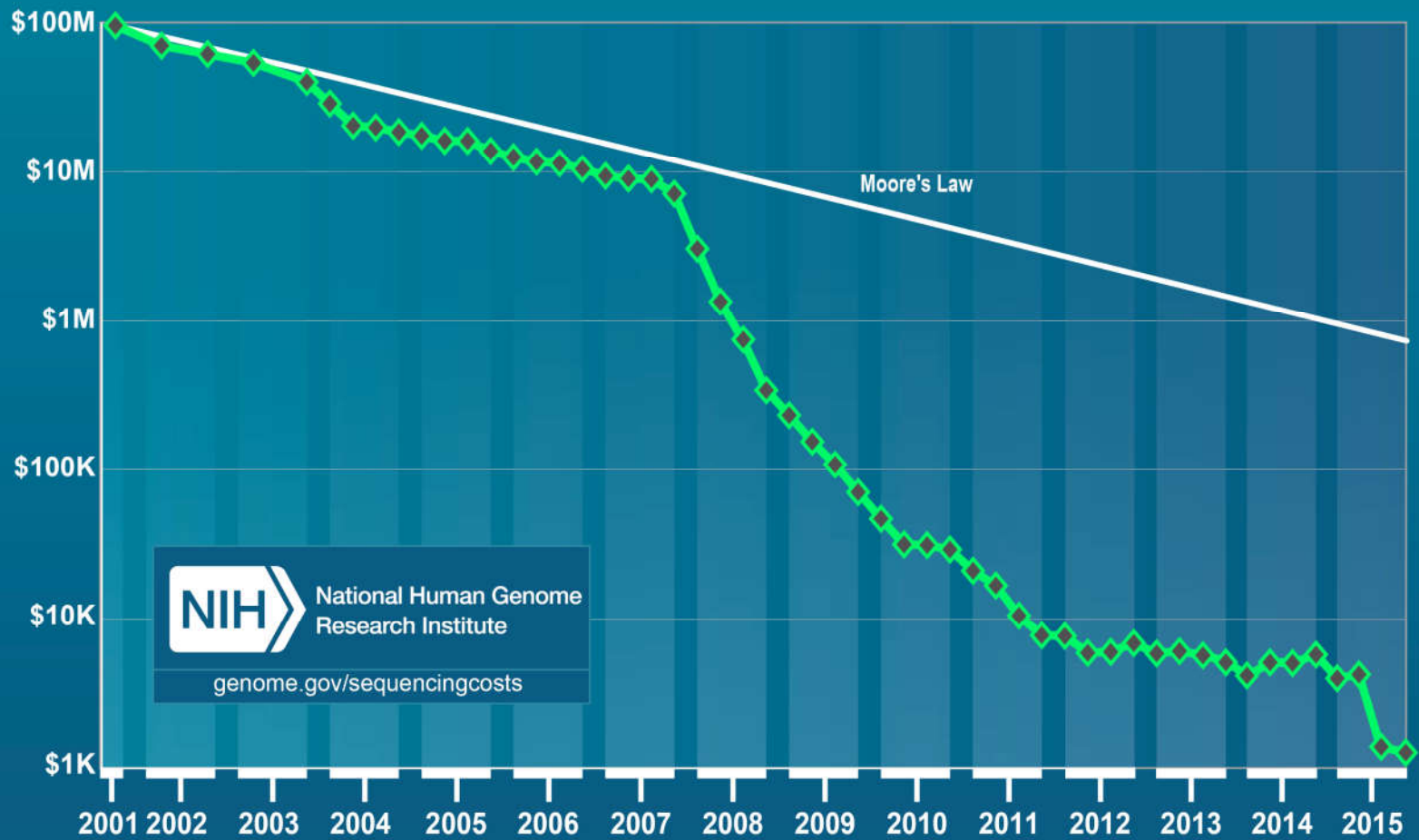
- Giải mã hệ gen người chỉ tốn \$1k (60 billion bases)!!



# Tiến bộ trong công cụ giải mã genome

		<b>Công suất</b>
Phương pháp thủ công	Sanger	0.01 Mb/day
Giải mã tự động, thế hệ thứ nhất	Slab Gel	1.6 Mb/day
	Capillary Array Electrophoresis	4Mb/day
Giải mã tự động, thế hệ thứ 2	454/Roche	750 Mb/day
	Illumina	5,000 Mb/day
	SOLiD	5,000 Mb/day
Giải mã tự động, thế hệ thứ 3	Helicos	5,000 Mb/day
	Pacific BioSciences	48,000 Mb/day

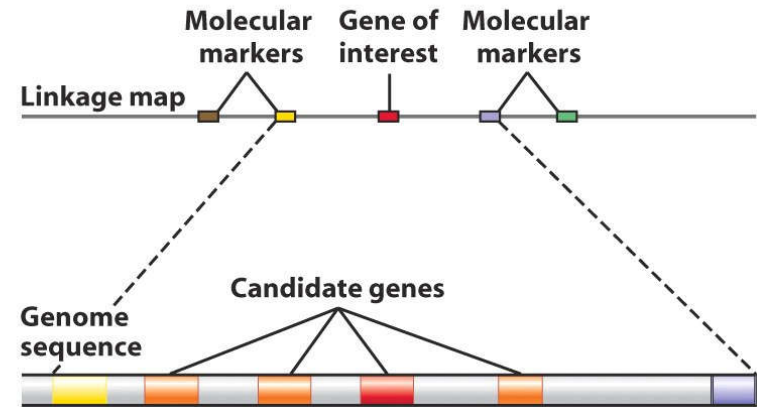
## Cost per Genome



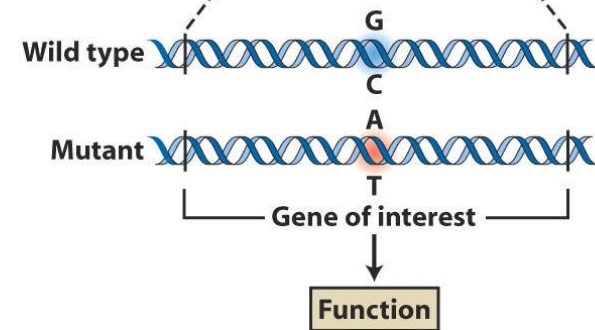
# Tại sao giải mã toàn bộ hệ gen?

- Thu được “các khối dữ liệu” giải thích cho việc tạo thành cơ thể sống
- Xác định được sự tương đồng và khác biệt giữa bộ gen tạo thành cơ thể sống này với cơ thể sống khác
- Am hiểu sâu về di truyền tiến hóa
- Nhanh chóng xác định các gen đích
- Ứng dụng hiệu quả trong y học, chọn, lai tạo giống

## 1 Localize gene of interest to a delimited region



## 2 Find exact location of gene of interest



## Các dự án lớn về giải mã hệ gen

**TABLE 15.6 Large-scale human sequencing projects (on-going and proposed).**

Project	URL	Goal
The 1000 Genomes Project	<a href="http://www.1000genomes.org/">http://www.1000genomes.org/</a>	Find human genetic variants having frequencies >1%
International Cancer Genome Consortium (ICGC)	<a href="http://www.icgc.org/">http://www.icgc.org/</a>	Catalog mutations in tumors from 50 cancer types
UK10K	<a href="http://www.sanger.ac.uk/about/press/2010/100624-uk10k.html">http://www.sanger.ac.uk/about/press/2010/100624-uk10k.html</a>	Sequence the genomes of 10,000 UK individuals
100,000 Genomes Project	<a href="http://www.genomicsengland.co.uk/">http://www.genomicsengland.co.uk/</a>	Sequence 100,000 individuals in the UK
Autism Genome 10K Project	<a href="http://autismgenome10k.org/">http://autismgenome10k.org/</a>	Sequence 10,000 autism-related genomes
Personal Genome Project	<a href="http://www.personalgenomes.org/">http://www.personalgenomes.org/</a>	Effort to sequence 100,000 human genomes

## Các dự án lớn về giải mã hệ gen

TABLE 15.7 Large-scale model organism sequencing projects (on-going and proposed).

Project	URL	Goal
1001 Genomes Project	<a href="http://www.1001genomes.org/">http://www.1001genomes.org/</a>	Find whole-genome sequence variation in 1001 strains of <i>Arabidopsis thaliana</i>
Genome 10K project	<a href="https://genome10k.soe.ucsc.edu/">https://genome10k.soe.ucsc.edu/</a>	Assemble sequences from 10,000 vertebrate species
Drosophila Genetic Reference Panel	<a href="http://dgrp2.gnets.ncsu.edu/">http://dgrp2.gnets.ncsu.edu/</a>	Sequence the genomes of 192 inbred lines from <i>Drosophila</i>
1000 Fungal Genomes Project	<a href="http://1000.fungalgenomes.org/home/">http://1000.fungalgenomes.org/home/</a>	Sequence 1000 fungal genomes
Mouse Genomes Project	<a href="http://www.sanger.ac.uk/resources/mouse/genomes/">http://www.sanger.ac.uk/resources/mouse/genomes/</a>	Sequence 17 mouse strains
Million Mutation Project	<a href="http://genome.sfu.ca/mmp/">http://genome.sfu.ca/mmp/</a>	<i>C. elegans</i>

# Các dự án giải mã genome lúa và ứng dụng

Materials	Depth of sequencing	SNPs	Research purposes	References
132 RILs of a super hybrid rice <i>Liang-You-Pei-Jiu</i>	>4x >36x for parents	171,847	Improving parental genome sequences Dissecting yield-associated loci	Gao et al. 2013
1083 cultivated rice* 446 wild rice	>1 ~ 50x	7,970,359	Domestication analysis of cultivated rice Identifying agronomic QTL	Huang et al. 2012a
40 cultivated rice 10 wild rice	> 15x	6,500,000	Identifying agronomic QTL Domestication analysis	Xu et al. 2012
950 cultivated rice	>1x	4,109,366	GWAS study of flowering time and grain yield traits	Huang et al. 2012b
517 rice landraces	>1x	3,625,200	GWAS study of 14 agronomic traits	Huang et al. 2010
150 RILs of Nipponbare/93-11	>20x	1,226,791	Large-scale gene discovery Identifying 49 QTLs for 14 agronomic traits	Huang et al. 2009 Wang et al. 2011
128 CSSLs of Nipponbare/93-11	>0.13x	7,680,000	QTL mapping for culm length High-throughput genotyping	Xu et al. 2010
5 cultivated rice	>58x	1,154,063	Genetic diverse analysis	Jeong et al. 2013
A restorer line 7302R 4 other cultivated rice	>13x	307,627	Genetic variation identification	Li et al. 2012
241 RILs of a hybrid rice Shanyou 63	> 0.06x	270,820	QTL detection for grains	Yu et al. 2011
40 RILs of Nortai/Hitomebore (bulked) 50 F <sub>2</sub> lines of Dunghan Shali/ Hitomebore	>6x	161,563	Rapid QTL mapping	Takagi et al. 2013
781 F <sub>2</sub> lines of R1128/Nipponbare	>16x	74,329	Genetic analysis for super hybrid rice	Duan et al. 2013
Koshihikari	15.7x	67,051	Evaluate the dynamics of the genome composition	Yamamoto et al. 2010

## Key allelic loci fine-mapped using the NGS method in rice

Trait	QTL	Chromosome	Mapping materials	Reference	Allelic loci
Tiller angle	qTA-9	9	Nipponbare/93-11	Wang et al. 2011	<i>TAC1</i> (Yu et al. 2007)
Plant height	qPH-1	1	Nipponbare/93-11	Wang et al. 2011	<i>SD1</i> (Sasaki et al. 2002)
Flag leaf width	qFLW-4	4	Nipponbare/93-11	Wang et al. 2011	<i>NAL1</i> (Qi et al. 2008)
Grain length	qGL-3	3	Nipponbare/93-11; PA64s/93-11; Zhenshan 97/Minghui 63	Wang et al. 2011; Gao et al. 2013; Yu et al. 2011	<i>GS3</i> (Fan et al. 2006)
Grain width	qGW-5	5	Nipponbare/93-11	Wang et al. 2011	<i>qSW5</i> (Shomura et al. 2008)
Heading date	qHD8	8	PA64s/93-11	Gao et al. 2013	<i>DTH8</i> (Wei et al. 2010)
Plant height	qPH5	5	PA64s/93-11	Gao et al. 2013	<i>EUI1</i> (Luo et al. 2006)
Plant height	qPH12	12	PA64s/93-11	Gao et al. 2013	<i>NRL1</i> (Hu et al. 2010)
Effective tiller number	qETN4	4	PA64s/93-11	Gao et al. 2013	<i>HTD1</i> (Zou et al. 2005)
Secondary panicle branch No.	qSPB1	1	PA64s/93-11	Gao et al. 2013	<i>LAX1</i> (Komatsu et al. 2001)
Seed set	qSS12	12	PA64s/93-11	Gao et al. 2013	<i>P/TMS12-1</i> (Zhou et al. 2012)
Hull color	Domesticaton sweeps	4	Natural population	Huang et al. 2012a	<i>Bh4</i> (Zhu et al. 2011)
Tiller angle	Domesticaton sweeps	7	Natural population	Huang et al. 2012a; Xu et al. 2012	<i>PROG1</i> (Jin et al. 2008; Tan et al. 2008)
Seed shattering	Domesticaton sweeps	4	Natural population	Huang et al. 2012a	<i>Sh4</i> (Li et al. 2006a)
Grain width	Domesticaton sweeps	5	Natural population	Huang et al. 2012a	<i>qSW5</i> (Shomura et al. 2008)
Leaf sheath color & apiculus color	Domesticaton sweeps	6	Natural population	Huang et al. 2012a	<i>OsC1</i> (Saitoh et al. 2004)
Seed shattering	Domesticaton sweeps	1	Natural japonica population	Huang et al. 2012a	<i>qSH1</i> (Konishi et al. 2006)

# Key allelic loci fine-mapped using the NGS method in rice

Grain quality	Domestication sweeps	6	Natural japonica population	Huang et al. 2012a	<i>Waxy</i> (Wang et al. 1995)
Pericarp color	Domestication sweeps	7	Natural japonica population	Huang et al. 2012a	<i>Rc</i> (Sweeney et al. 2006)
Grain width	gw5, kgw5	5	Zhenshan 97/Minghui 63	Yu et al. 2011	<i>GW5/qSW5</i> (Shomura et al. 2008)
Pigmentation	qPIG6	6	Zhenshan 97/Minghui 63	Yu et al. 2011	<i>OsC1</i> (Saitoh et al. 2004)
Gelatinization temperature	qGT6	6	517 rice landraces	Huang et al. 2010	<i>ALK</i> (Gao et al. 2003)
Plant height	qPH1-3	1	R1128/Nipponbare	Duan et al. 2013	<i>Sd1</i> (Sasaki et al. 2002)
Heading date	qPBN6-2	6	R1128/Nipponbare	Duan et al. 2013	<i>Hd1</i> (Yano et al. 2000)
Early heading date	qPBN10-1	10	R1128/Nipponbare	Duan et al. 2013	<i>Ehd1</i> (Doi et al. 2004)
Grain number 1	qGN1-1	1	R1128/Nipponbare	Duan et al. 2013	<i>Gn1</i> (Ashikari et al. 2005)
Ideal plant architecture	qPL8-1	8	R1128/Nipponbare	Duan et al. 2013	<i>IPA1</i> (Jiao et al. 2010)





**Công cụ Genome trong chọn tạo  
giống cây trồng**

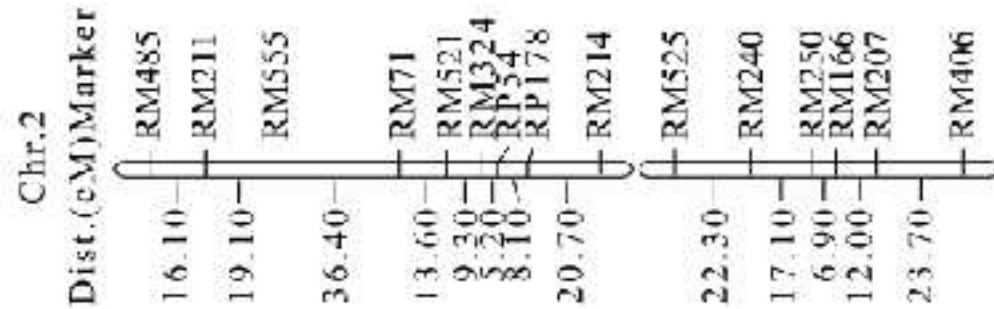
# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

## 1. Xây dựng bộ chỉ thị phân tử với mật độ dày đặc bao phủ toàn bộ các NST

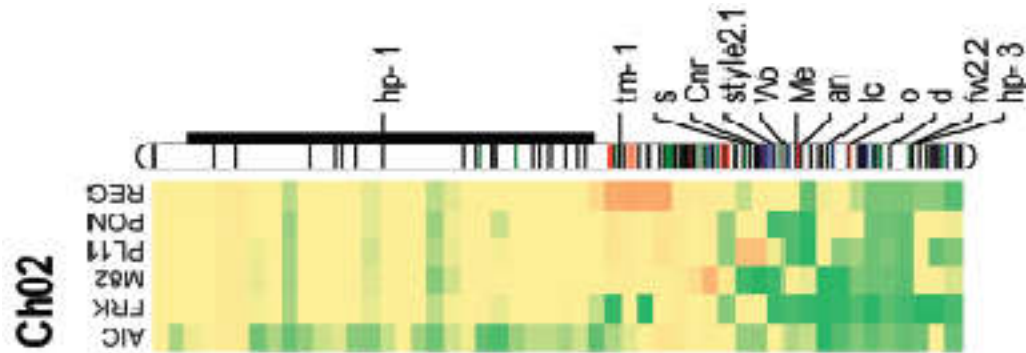
### + Tăng mật độ các chỉ thị:

- SSRs-simple sequence repeat (trình tự lặp đơn giản), SNPs - single nucleotide polymorphism , InDels (các đoạn chèn/mất)
- Ở lúa, trước khi bộ gen được giải mã và công bố, có hơn 2200 SSRs (2002) dựa vào các clone PAC/BAC
- Sau khi hoàn tất bộ gen lúa Nipponbare, có thêm 18.828 SSRs (Sasaki, 2004)

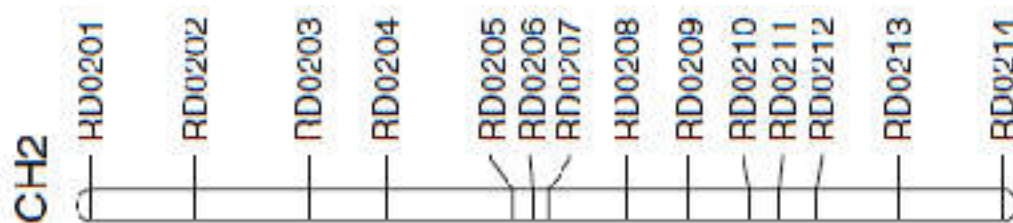
# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng



**SSR map**, reference Nipponbare - 9311 population (Zhang Qi-Jun et al. 2006)



**SNP map** in tomato, reference SL2.40 (Shirasawa et al. 2013)

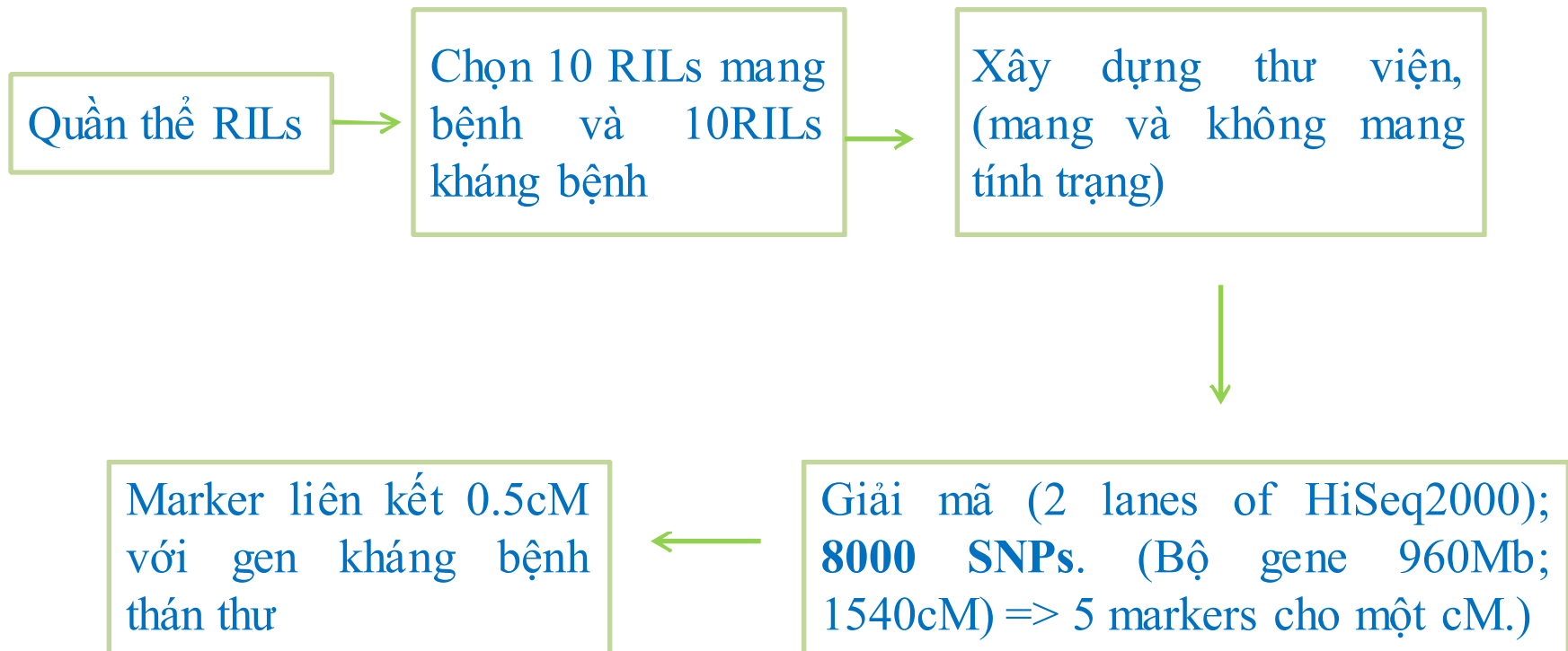


**InDels map** in rice, with genome sequence (Dong-Hong Wu et al. 2013)

# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

## + Tìm chỉ thị liên kết

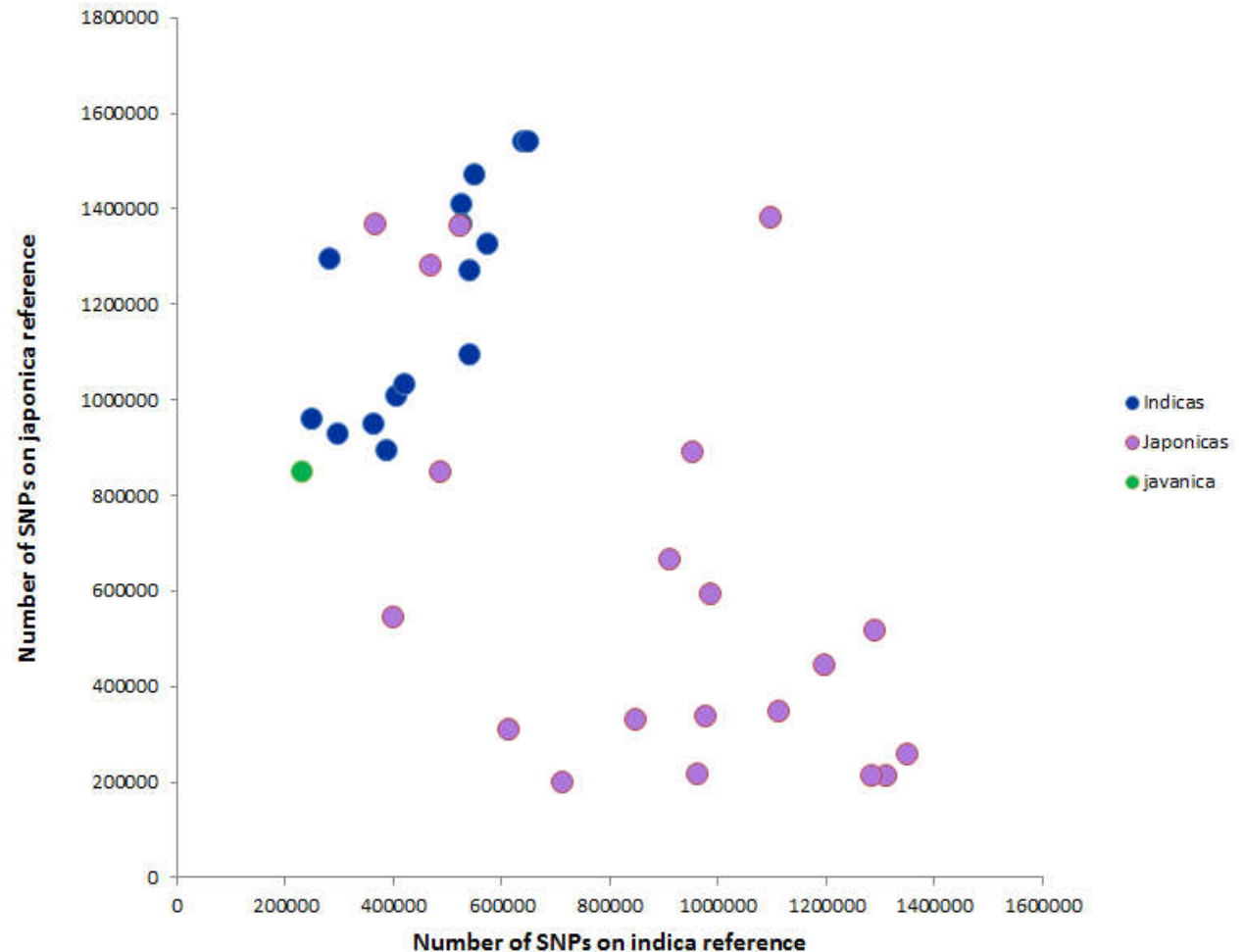
Ví dụ: tìm chỉ thị liên kết với gen kháng bệnh thán thư ở Cây họ đậu  
Lupinus (Yang *et al.*, BMC Genomics; 2012)



# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

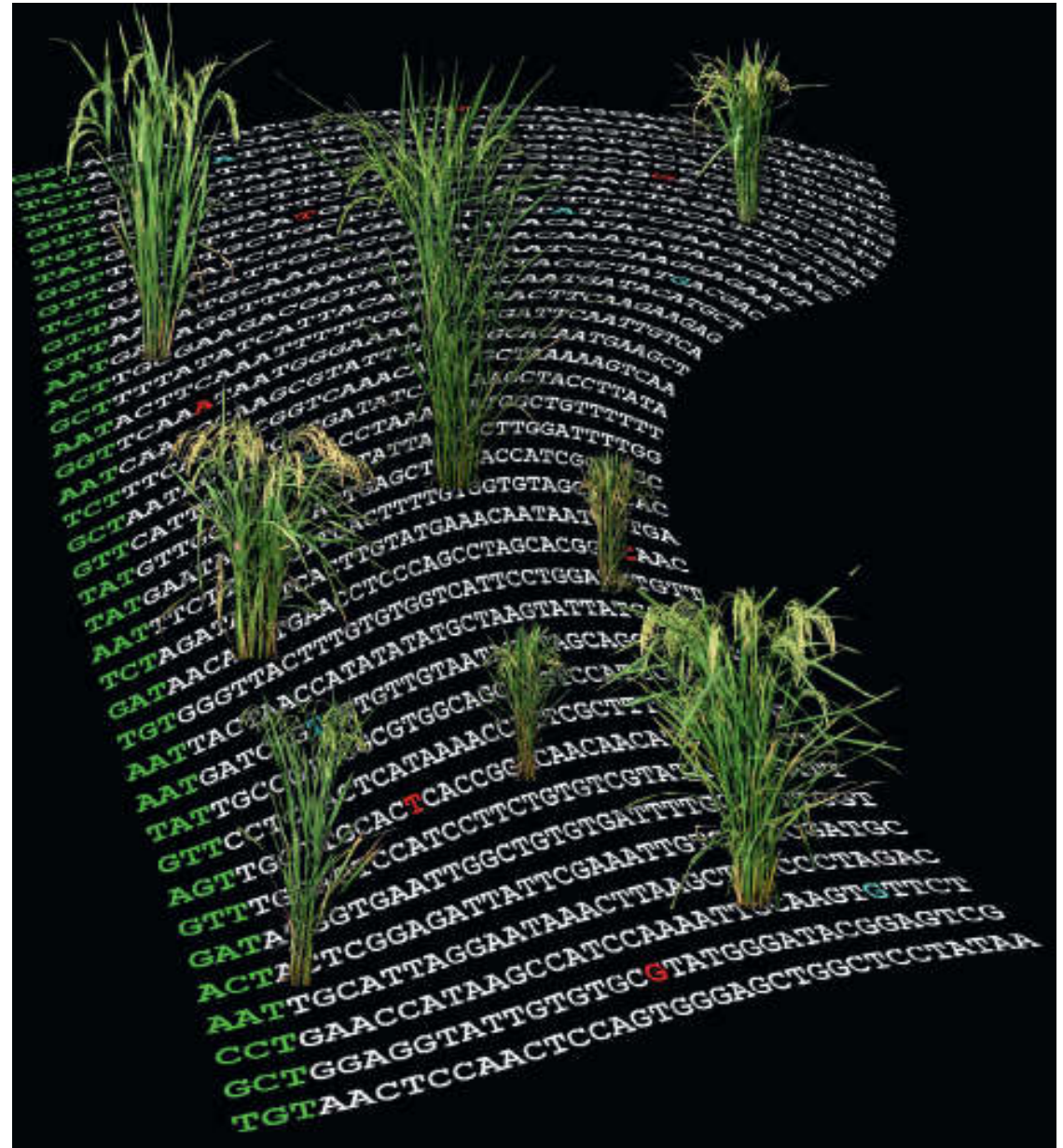
## 2. Lập bản đồ SNPs, đánh giá đa dạng di truyền

Biểu đồ so sánh số lượng SNPs của các giống lúa Việt Nam với dòng indica 93-11 và dòng Nipponbare japonica tham chiếu



# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

- Ở lúa, xác định SNPs dựa trên so sánh các trình tự của tổ hợp lúa khác nhau với trình tự hệ gen quy chiếu.
- Từ đó, đánh giá đa dạng kiểu hình, tính trạng nông học dựa trên khác biệt kiểu gen (Bin Han và Xuehui Huang, 2013)



# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

## 3. Ứng dụng để tìm các candidate genes

- Tầm soát các đoạn trình tự tương đồng (homologous gen) từ trình tự thô của toàn bộ genome lúa để xác định các candidate gen

-Xác định candidate gen bằng nsSNP (non synonymous SNP) sử dụng phương pháp chọn lọc các biến dị chung (common variant – CV)

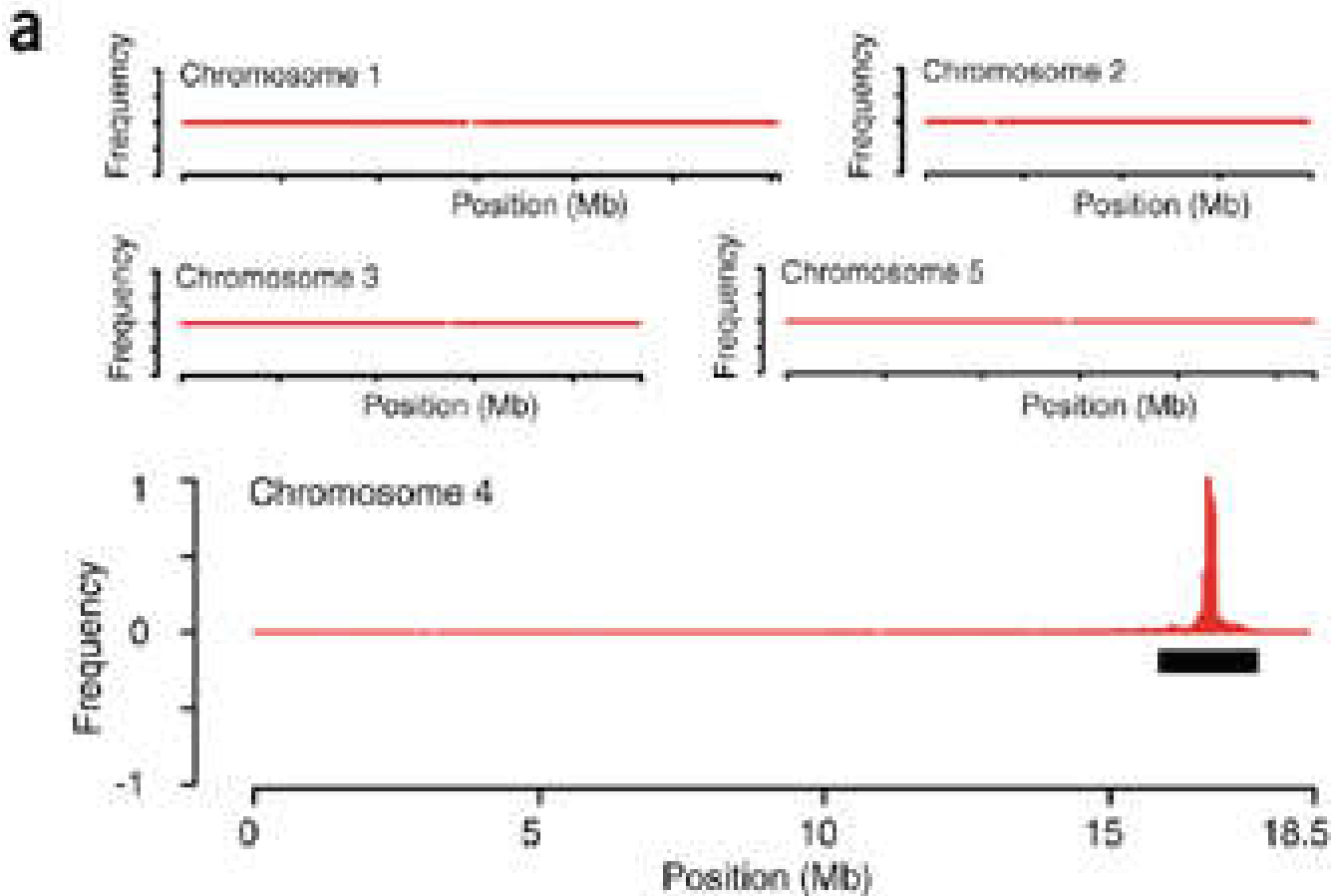
-Phương pháp nhận diện các trình tự NBS-LRR (nucleotide binding site-leucine rich repeat) cơ bản trong genome lúa

**Ví dụ:** Tìm ra candidate gene mới liên kết với tính miễn cảm với artemisinin ở loài kí sinh trùng gây sốt rét (*Plasodium falciparum*) ở Kenya bằng phân tích 75.000 SNPs chọn lọc từ giải trình tự hệ gen (Borrmann et al. 2013).

# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

## 4. Tìm đột biến bằng giải mã toàn bộ hệ gene

Giải mã một nhóm 20 cá thể mang kiểu hình đột biến và 20 nhóm cá thể không mang kiểu hình đột biến trong một quần thể F<sub>2</sub>. Từ đó xác định được đột biến (chỉ xuất hiện ở 20 nhóm cá thể mang kiểu hình đột biến).



(Schneeberger et al., 2009)



# Ứng dụng của công cụ genome trong chọn tạo giống cây trồng

## 5. Chọn lọc hệ gene (genomic selection); GWAS và lập bản đồ tính trạng số lượng (QTL mapping)

+ **Genomic selections:** thúc đẩy nhanh việc xác định kiểu gene của các cá thể trong đời F2/F3 trên nhiều locus một lúc để cùng một lúc chọn lọc/qui nạp được nhiều gene tốt/QTL tốt.

Ví dụ, ở lúa, lập bản đồ QTL của 14 tính trạng nông học cùng một lúc, chẳng hạn như hàm lượng amylose, chịu hạn, độ dài hạt, trọng lượng hạt...(Xuehui Huang et al. 2013 )

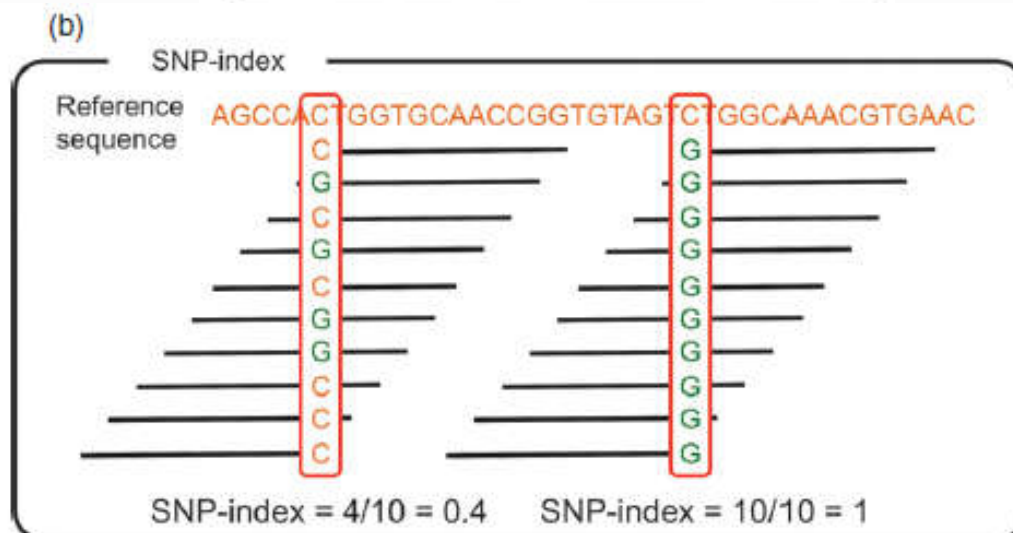
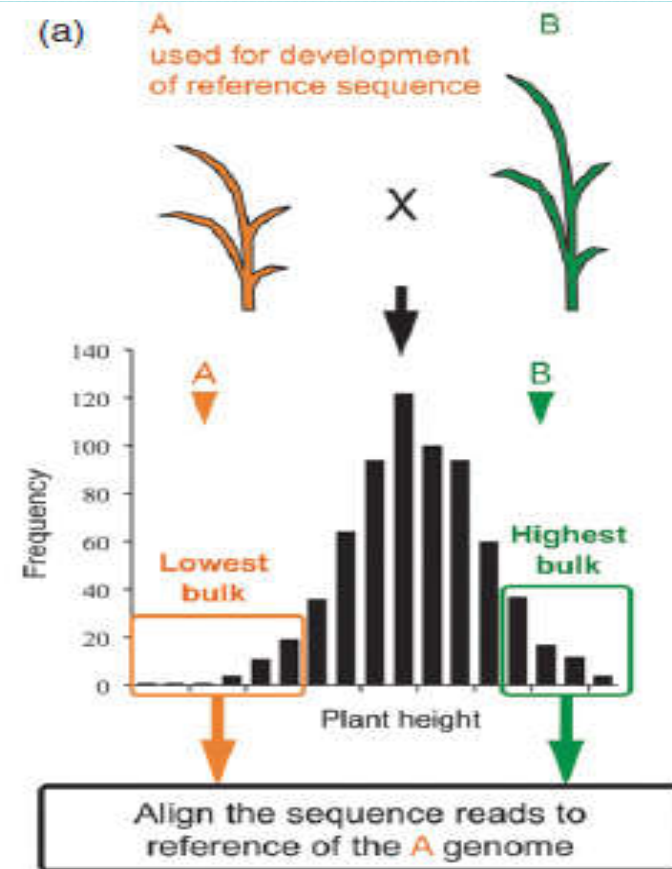
Trait	Chr.	Position (IRGSP4)	Major allele	Minor allele	Minor allele freq	$-\log_{10}P$ (Simple model)	$-\log_{10}P$ (Compressed MLM)
Amylose content	3	18,705,982	C	T	0.06	16.81	1.83
Amylose content	6	1,757,040 <sup>b</sup>	C	G	0.14	67.75	25.30 (1,770,929 <sup>a</sup> )
Amylose content	6	6,189,558	A	T	0.11	34.76	7.52
Amylose content	6	6,709,537	C	T	0.19	35.45	11.13
Amylose content	12	10,993,688	G	T	0.06	15.93	1.78
Apiculus Color	6	5,335,519 <sup>b</sup>	A	G	0.33	48.44	26.25 (5,335,519 <sup>a</sup> )
Apiculus Color	6	7,681,502	G	A	0.32	17.31	8.48
Apiculus Color	12	460,120	T	C	0.22	9.92	1.60
Drought tolerance	1	5,536,395	G	T	0.11	9.97	6.39
Drought tolerance	2	1,489,158	T	C	0.12	9.32	4.57
Drought tolerance	5	2,275,357	A	C	0.06	14.24	7.61
Drought tolerance	6	28,243,628	C	T	0.09	11.98	8.47
Drought tolerance	11	21,161,361	G	C	0.08	15.28	11.07



## 5. Chọn lọc hệ gene (genomic selection); GWAS và lập bản đồ tính trạng số lượng (QTL mapping)

+ **QTL mapping**: giúp xác định kiểu gene của các cá thể trong quần thể để lập bản đồ tính trạng số lượng.

Ví dụ, lập bản đồ QTL-seq ở lúa dựa vào trình tự DNA toàn hệ gen từ 2 quần thể, mỗi quần thể gồm 20-50 cá thể có tính trạng đối lập (Hiroki Takagi et al. 2013)



## 5. Chọn lọc hệ gene (genomic selection); GWAS và lập bản đồ tính trạng số lượng (QTL mapping)

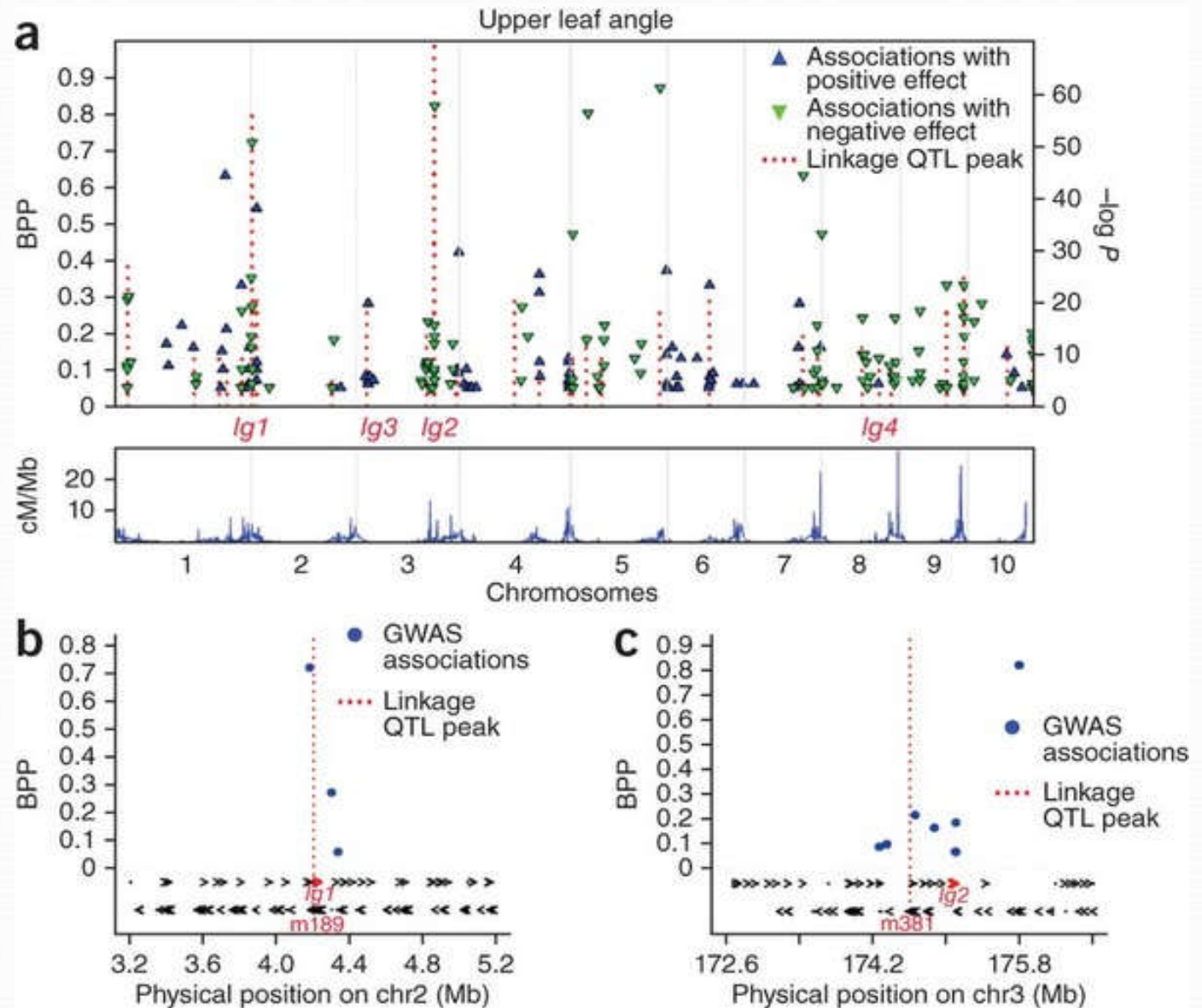
+ **GWAS (Genome-wide association study)**: Đối với những đối tượng không thể lai tạo để tạo quần thể nội phối hoặc tạo quần thể mất hàng chục năm (cây lâu niên, cây khó thu hạt); việc giải mã hàng nghìn cá thể kết hợp với đánh giá kiểu hình có thể xây dựng mối liên kết giữa kiểu gene với kiểu hình.

Ví dụ, GWAS trên cây táo (*Malus x domestica*) để phát hiện họ gen GH3 (Gretchen Hagen3) duy trì sự cân bằng hormon (IAA, SA and JA). Nhờ giải trình tự hệ gen, xác định gen với vai trò đáp ứng abiotic/biotic stress (Yuan, 2013)



## 5. Chọn lọc hệ gene (genomic selection); GWAS và lập bản đồ tính trạng số lượng (QTL mapping)

Ví dụ, GWAS ở ngô tìm thấy đột biến gene *ligules 1, 2* (*lg1; lg2*) gây ra góc lá đứng sau khi áp dụng giải mã với 5000 cá thể. (Tian et al., 2011)

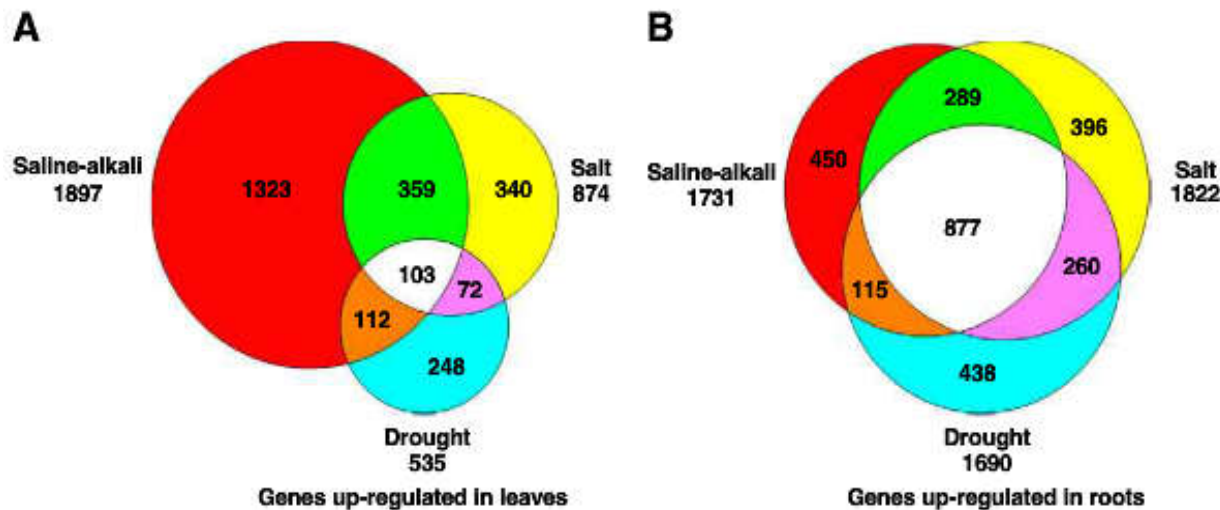


## 6. Nghiên cứu biểu hiện của gene RNA-seq và E-QTLs (gene expression QTLs)

- NGS ra đời đã thay thế microarray.

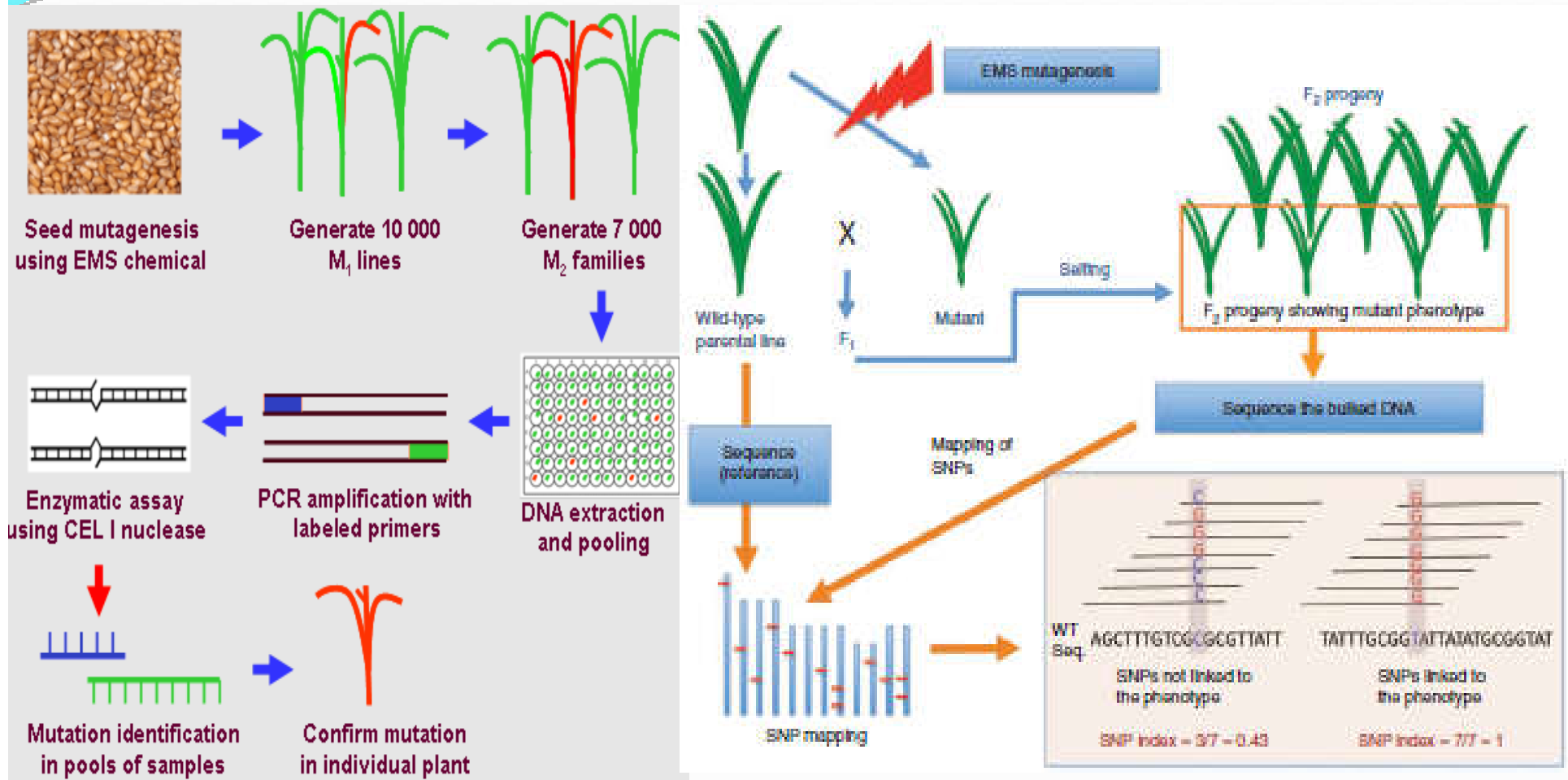
- NGS (next gene sequencing) có khả năng đọc và đếm số reads. Nên thực hiện giải mã dùng cDNA library thì có thể xác định xem một gene được biểu hiện nhiều hay ít dựa vào số lần gene đó được đọc giải mã.

**Ví dụ:** về biểu hiện gen của lá và rễ đậu tương trong điều kiện mặn, kiềm hoặc hạn hán bằng giải trình tự Illumina



Số ở ngoài vòng tròn chỉ số gen biểu hiện tính trạng ở lá (A) và rễ (B) dưới mỗi điều kiện. Số trong vòng tròn chỉ các gen đặc hiệu với stress. Số ở 2 hoặc hơn 2 vòng tròn giao nhau chỉ gen đa tính trạng (Fan et al., 2013)

# 7. Phát hiện nhanh các đột biến được tạo ra



Phương pháp truyền thống  
TILLING

Phương pháp mới nhờ trình tự genome  
(Abe et al. 2011)

## TÌNH HÌNH GIẢI MÃ GENOME TẠI VIỆT NAM

---

- **Đề tài KC04.20-11/15:** Giải trình tự hệ gen loài vi tảo biển dinh dưỡng của Việt Nam *Schyzochytrium Mangrovei* PQ6
  - Thời gian: 1/1/2014-1/12/2015
  - Kích thước: 4,5Mb
  - Mục đích: xác định được trình tự hệ gen toàn phần loài vi tảo biển dị dưỡng của Việt Nam PQ6; phân tích chức năng và các thông tin liên quan đến trình tự gen đã được giải mã phục vụ cho nghiên cứu cơ bản và ứng dụng

# TÌNH HÌNH GIẢI MÃ GENOME TẠI VIỆT NAM

---

- **Chương trình giải mã genome người:**

**Đề tài:** Giải trình tự và xây dựng hoàn chỉnh hệ gen người Việt Nam đầu tiên làm “trình tự tham chiếu” và bước đầu phân tích nhân chủng học tiến hóa người Việt Nam

- Thời gian: 01/07/2015 - 01/07/2018

- Mục đích:

- + Giải trình tự toàn bộ hệ gen của một số gia đình (bố, mẹ và con) để lựa chọn trình tự tham chiếu (reference sequence) sơ bộ của người Việt Nam.

- + Giải trình tự toàn bộ hệ gen ty thể và nghiên cứu đa hình nucleotide đơn trên vùng không trao đổi chéo của nhiễm sắc thể Y của >300 cá thể thuộc 10 dân tộc của 5 nhóm ngôn ngữ sống tại Việt Nam qua đó xác định nguồn gốc các dân tộc, quan hệ phát sinh và nhân chủng học tiến hóa người Việt Nam.





## TÌNH HÌNH GIẢI MÃ GENOME VÀ ĐỊNH HƯỚNG ỨNG DỤNG TẠI VIỆT NAM

---

- **Đề tài giải mã transcriptome sâm Ngọc Linh:** Giải trình tự và phân tích hệ gen phiên mã (transcriptome) ở Sâm Ngọc Linh (*Panax vietnamensis* Ha et Grushv.).
  - Thời gian 2017-2019
  - Mục đích: Giải trình tự và phân tích được hệ gen phiên mã ở Sâm Ngọc Linh; Xây dựng được cơ sở dữ liệu hệ gen phiên mã của Sâm Ngọc Linh; Phát hiện được các nhóm gen tham gia chuỗi sinh tổng hợp ginsenoside của Sâm Ngọc Linh

# TÌNH HÌNH GIẢI MÃ GENOME LÚA VÀ ĐỊNH HƯỚNG-KHAI THÁC ỨNG DỤNG TẠI AGI

- **Đề tài giải mã genome lúa (pha I)**

- Thời gian: 2011-2013
- Số lượng: 36 giống, 20-60x

- **Đề tài giải mã genome lúa (pha II)**

- Thời gian: 2014-2017
- Số lượng: ~ 600 giống, 2-6x

- **Mục đích:**

- + Giải trình tự, xây dựng cơ sở dữ liệu genome của các dòng/giống lúa của Việt Nam phục vụ công tác nghiên cứu, khai thác, sử dụng trong các chương trình chọn tạo giống;
- + Đào tạo cán bộ chuyên sâu trong lĩnh vực giải mã genome, nghiên cứu genome học và ứng dụng bioinformatic trong quản lý, khai thác trình tự genome phục vụ chọn tạo giống.



# KẾT QUẢ NGHIÊN CỨU

## 1. Lập bản đồ SNPs, đánh giá đa dạng di truyền tập đoàn nguồn gen lúa bản địa

### + Lập bản đồ SNPs (36 giống pha I và 300 giống pha II)

- SNPs - single nucleotide polymorphism ,
- InDels (các đoạn chèn/mất)

### + Phân tích, đánh giá đa dạng di truyền

- Phân tích hệ gen của 5 giống lúa kháng đạo ôn
- Bài báo:

“Whole-Genome Characteristics and Polymorphic Analysis of Vietnamese Rice Landraces as a Comprehensive Information Resource for Marker-Assisted Selection. International Journal of Genomics Volume 2017 (2017), Article ID 9272363, 11 pages, SCIE, IF=1,89



# KẾT QUẢ NGHIÊN CỨU

---

## 2. Tìm các candidate genes

+ Tìm các candidate gene (kháng bạc lá, đạo ôn, rầy nâu, chịu mặn, chịu hạn)

# Các candidate gen có trong các giống lúa bản địa của Việt nam

	Tên	Phân loại	Đặc tính	Bad h2	Xa13	Xa21	Bph 14	Pi7	Pi21	pikp	Pita	Dreb 2a	Dreb 1a	Dreb 1b	Dreb 1d	Sall	Rss s1	SRW D4
1	Tam xoan Bac Ninh	J	Thơm	x								x						
2	Tam xoan Hai Hau	J	Thơm	x								x		x				
3	Te Nuong	J	Thơm	x		x	x					x		x		x	x	
4	Nang thom cho dao	J	Thơm	x														
5	Thom Lai		I	Thơm	x						x	x						
6	Nep man		I	Mặn														
7	Chiem do		I	Mặn		x						x						x
8	Lua Ngoi	J	Mặn															
9	Mot bui do	J	Mặn		x		x				x	x		x	x	x		x
10	Nang co do 2		I	Mặn														
11	Ble te lo	J	Đạo ôn		x			x	x	x	x		x	x				x
12	Chiem nho BN2		I	Đạo ôn				x										
13	Nep lun		I	Đạo ôn	x													
14	Khau mac buoc	J	Đạo ôn			x					x			x		x	x	x
15	OM6377		I	Đạo ôn	x							x						x
16	Chan thom	J	Rây nâu	x	x						x							x
17	Xuong ga		I	Rây nâu		x												x
18	Khau giang	J	Rây nâu								x							x
19	Coi ba dat	J	Rây nâu															
20	OM5629	J	Rây nâu		x								x					x
21	Nep bo hong HD		I	Hạn		x			x			x						
22	Tan ngan	J	Hạn									x		x				
23	Ba cho K'te	J	Hạn															
24	Blao sinh sai	J	Hạn								x	x						x
25	Nang quot bien		Hạn		x							x						x
26	Tep Thai Binh		I	Bạc lá														
27	Khau dien lu	J	Bạc lá											x				
28	Nep meo nuong	J	Bạc lá			x			x			x						x
29	Toc lun		I	Bạc lá														
30	Hom rau	J	Bạc lá		x	x												
32	Nep ong tao	J	Thơm															
33	OM3536		I	Thơm														x
34	Khau Lien	J	Rây nâu											x				x
36	Lua goc do		I	Đạo ôn				x										
37	Chiem da		I	Mặn														
39	IS1.2		I	Rây nâu				x				x						x

# KẾT QUẢ NGHIÊN CỨU

## 3. Lai tạo, đánh giá khả năng kháng bệnh và chống chịu của các candidate gene có trong các giống lúa bản địa của Việt Nam (thế hệ BC<sub>3</sub>F<sub>3</sub>)

### 3.1. Gen kháng Rầy nâu, chịu hạn, chịu mặn

TT	Gen kháng	Tên tổ hợp lai	Tên cặp mồi	Kích thước	Số lượng cá thể BC <sub>3</sub> F <sub>3</sub> mang gen kháng
<b>A</b>	Tổ hợp lai có các con lai mang 1 gen				
<b>1-Rầy nâu</b>	Bph14	TSL1/ OM5629	Bph14 add27	176bp	04
		BC15/ OM5629	Bph14 add27	176bp	07
		TSL1/ OM6377	Bph14 add27	176bp	06
		Thủ đô 1/ Chân thơm	Bph14 add27	176bp	05
		Bắc thơm số 7/ Lúa ngoi	Bph14 add27	176bp	05
<b>2-Chịu Mặn</b>	Salt1	Bắc thơm số 7/ OM5629	Salt1 add6	113bp	09
	SRWD5	Bắc thơm số 7/ Coi ba đất	SRWD5 dell200	225bp	09
<b>3-Chịu Hạn</b>	Dreb1C	TSL1/OM6377	Dreb1C del6	177bp	06
<b>B</b>	Tổ hợp lai có các con lai mang 2 gen				
<b>1-Rầy nâu+Chịu hạn</b>	Bph14 + Dreb1C	TSL1/OM6377	Bph14 add27	176bp	04
			Dreb1C del6	177bp	

# KẾT QUẢ NGHIÊN CỨU

## 3. 2. Gen kháng Bạc lá

TT	Gen kháng	Tên tổ hợp lai	Tên cặp mồi	Kích thước	Số lượng con lai mang gen kháng
<b>A</b>	<b>Tổ hợp lai có các con lai mang 1 gen</b>				
1	Xa3	Bắc thơm số 7/Tan ngàn	Xa3de40	178bp	02
2	Xa4	Sơn Lâm 1/Chấn thơm	MP1-MP2	150bp	03
3	xa5	OM6976/Chiêm nhỡ Bắc Ninh 2	xa5add35	179bp	06
		An dân 11/Chiêm nhỡ Bắc Ninh 2	xa5add35	179bp	04
		Thủ đô 1/Chiêm nhỡ Bắc Ninh 2	xa5add35	179bp	04
		DT39 Quế Lâm/OM6377	xa5add35	179bp	03
		Q1-8-1/Chấn thơm	xa5add35	179bp	04
<b>B</b>	<b>Tổ hợp lai có các con lai mang 2 gen</b>				
1	xa5	Thủ đô 1/Chấn thơm	xa5add35	179bp	03
	Xa4		MP1-MP2	150bp	
2	xa5	DT39 Quế Lâm/Chấn thơm	xa5add35	179bp	03
	Xa4		MP1-MP2	150bp	
3	Xa7	DT39 Quế Lâm/Chấn thơm	P3	300bp	02
	Xa4		MP1-MP2	150bp	
<b>C</b>	<b>Tổ hợp lai có các con lai mang 3 gen</b>				
1	xa5	An Dân 11/Hom râu	xa5add35	179bp	05
	Xa7		P3	300bp	
	xa13		Xa13add4	95bp	

# KẾT QUẢ NGHIÊN CỨU

## 3. 3. Gen kháng Đạo ôn

TT	Gen kháng	Tên tổ hợp lai	Tên cặp mồi	Kích thước	Số lượng con lai mang gen kháng
<b>A</b>	<b>Tổ hợp lai có các con lai mang 1 gen</b>				
<b>1</b>	Pita	Thủ đô 1/Chiêm nhỡ Bắc Ninh 2	Pita-Taq1	140bp	06
		Bắc thơm số 7/Lúa ngoi	Pita-Taq1	140bp	05
		An Dân 11/Chiêm nhỡ Bắc Ninh 2	Pita-Taq1	140bp	05
<b>2</b>	pikp	RVT/OM5629	Pikpdell6	174bp	06
		BC15/OM5629	Pikpdell6	174bp	07
		Thủ đô 1/Tốc lùn	Pikpdell6	174bp	06
		Jasmine/Tốc lùn	Pikpdell6	174bp	06
		An Dân 11/Nếp lùn	Pikpdell6	174bp	04
<b>B</b>	<b>Tổ hợp lai có các con lai mang 2 gen</b>				
<b>1</b>	Pita/pikp	Bắc thơm số 7/ OM5629	Pita-Taq1	140bp	04
			Pikpdell6	174bp	



# ĐỊNH HƯỚNG NGHIÊN CỨU

## 1. Tiếp tục lập bản đồ SNPs, đánh giá đa dạng di truyền nguồn gen lúa bản địa của Việt Nam

- + Lập bản đồ SNPs (của 300 giống)
- + Khai thác đa dạng di truyền của 600 giống mới giải mã
- + Dựa trên kết quả phân tích, đánh giá đa dạng di truyền 8.000-10.000 nguồn gen lúa bản địa của Việt Nam (dự án hợp tác với UK)
- + Viết bài

## 2. Tiếp tục tìm các candidate genes

- + Tìm các candidate gene (kháng bạc lá, đạo ôn, rầy nâu, chịu mặn, chịu hạn...)
- + Thiết kế các marker phân tử xác định các candidate gene
- + Lai tạo, đánh giá khả năng kháng bệnh và chống chịu của các candidate gene có trong các giống lúa bản địa của Việt Nam phục vụ công tác chọn tạo giống

# ĐỊNH HƯỚNG NGHIÊN CỨU

## 3. Phát hiện nhanh các đột biến được tạo ra

- + Đối sánh giữa trình tự giống gốc và giống đột biến, tìm các gen đột biến (Gen hạt to của giống Bắc thơm đột biến; Gen qui định hàm lượng amylose, TGST..)
- + Lai tạo, nghiên cứu chức năng, sự biểu hiện gen, ứng dụng trong chọn giống

## 4. GWAS và lập bản đồ tính trạng số lượng (QTL mapping)

- + Đánh giá kiểu hình (chịu mặn, hạn, kháng sâu bệnh...)
- + GWAS (lập bản đồ xác định các tính trạng nông học quan trọng phục vụ công tác chọn tạo giống)

## 5. Xây dựng nhóm tin sinh và nâng cấp PTN tin sinh học

- + Thành lập nhóm tin sinh học của Viện, phối hợp các chuyên gia trong và ngoài nước để khai thác cơ sở dữ liệu đã giải mã
- + Nâng cấp phòng thí nghiệm tin sinh học (máy chủ, máy trạm, hệ thống lưu trữ dữ liệu...)



**THANK YOU**

THANK YOU