Marker-Assisted Approaches, Candidate Gene Validation and Decision Support

Wang Lihao

Chinese Academy of Agricultural Sciences
No.12 Zhongguncun south street Beijing ,100081
Email: wanglihao@caas.cn

Cibinong, Indonesia, 2026.4.14

Contents

01. Introduction: Challenges & Rise of Molecular Breeding

Achievements and Challenges of Traditional Breeding

Green Revolution

Application of semi-dwarf genes significantly increased yields of rice, wheat, and other crops, solving food shortages.

Variety Improvement

Development of high-yield, high-quality, and disease-resistant varieties effectively ensured global food security.

Adaptive Breeding

Breeding varieties adapted to different ecological zones and farming systems expanded crop cultivation areas.

IR8 was developed by crossing DGWG with Peta

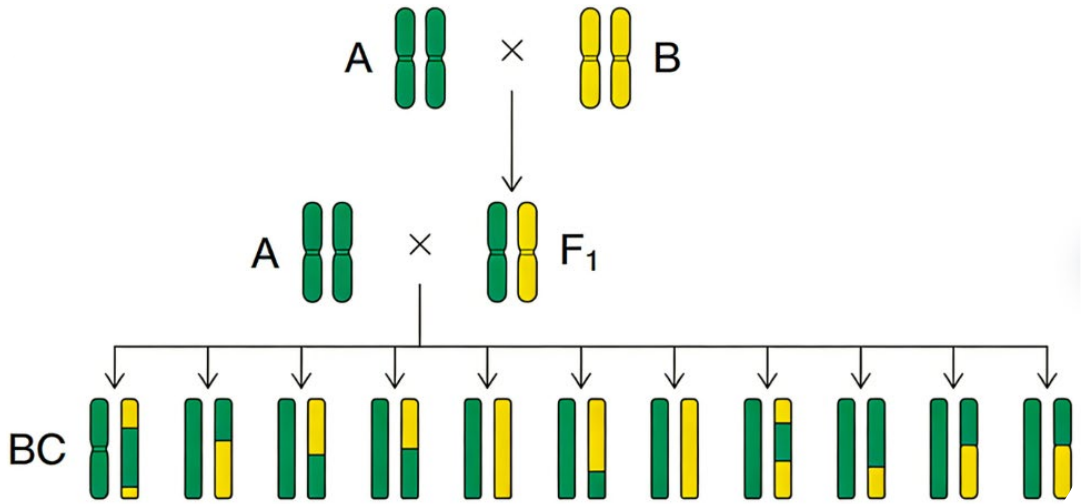
DGWG is a short, sturdy semi-dwarf landrace. Peta is a tall, highly tillering tall-stature landrace, originating from Indonesia.

Conclusion: Traditional breeding laid a solid foundation, but its limitations highlight the need for integration with molecular breeding to meet future food security demands.

Challenges of Traditional Breeding

Long Breeding Cycle

From crossbreeding to stable varieties, it requires **8-10 generations**, and rice crossbreeding takes an average of **more than 10 years**.



Low Efficiency

Traditional breeding relies heavily on phenotypic observation, resulting in the elimination of many non-target individuals, leading to **severe waste of human and material resources** and low overall screening efficiency.

Challenges of Traditional Breeding

Phenotype: Gene-Environment Interaction

Appearance is influenced by both genotype and environment (light, temperature).

Limited Selection Accuracy

Hard to determine if phenotype comes from genes or environment, affecting precision.

The same variety shows huge yield differences between arid and humid regions; field phenotypes alone are insufficient to assess true potential.

Comparison of Environmental Effects on Crop Phenotypes

Challenges of Traditional Breeding

Polygenic Control

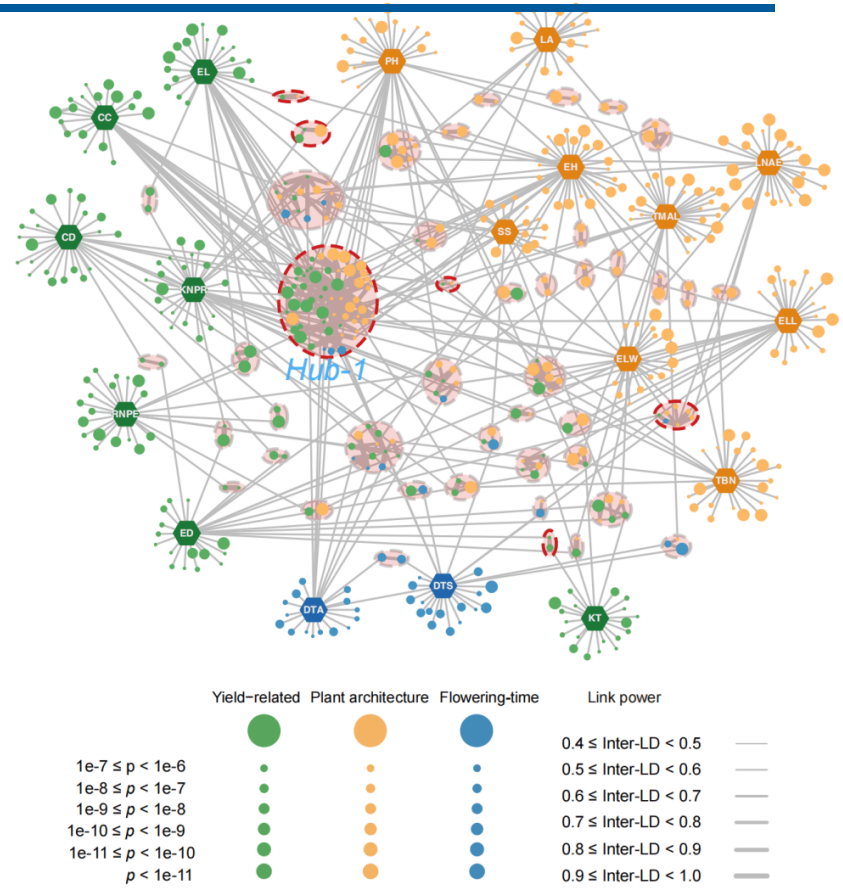
Important traits such as yield and quality are controlled by multiple genes, with complex genetic bases, making precise regulation difficult.

Gene Interactions

Complex interactions like additive, dominant, and epistatic effects make genetic rules hard to parse and improvement directions unpredictable.

Linkage Drag

Target genes are often tightly linked to unfavorable ones, making complete separation through traditional backcrossing difficult and hindering the aggregation of beneficial traits.



Genetic interaction network among 18 agronomic traits in maize

From Empirical Selection to Scientific Design

Experience-dependent,

Phenotypic Selection

Long cycle, Low efficiency

Dimensional Shift

From "Visible" Phenotype to "Invisible" Genotype

Paradigm Shift

From "Accidental" Excellent Variation to "Inevitable" Gene Pyramiding

Based on Genotypic Selection

Precise, Efficient, Designable

The Rise and Advantages of Molecular Breeding

Precise Selection

Direct selection based on genotype, unaffected by environment, high accuracy. Can directly track target gene transmission.

Cycle Shortening

Early selection possible, screening at seedling stage, significantly reducing breeding years.

Gene Pyramiding

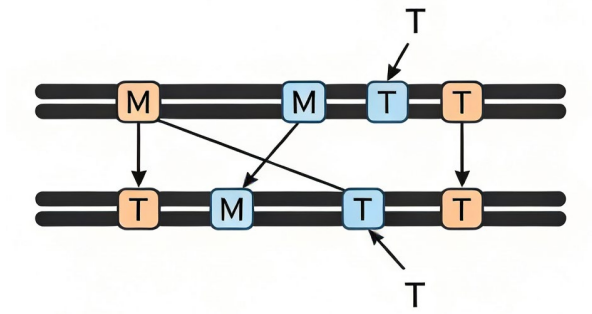
Simultaneous selection and pyramiding of multiple target trait genes to develop varieties with excellent comprehensive traits.

Resource Efficiency

Fully exploit and utilize excellent alleles in germplasm resources to broaden the genetic basis of breeding.

02. Overview: Marker-Assisted Selection (MAS)

Definition and Basic Principles of MAS



Linkage between Molecular Marker (M) and Target Gene (T)

Core Definition

Using **molecular markers** tightly linked to target traits for **indirect selection** of breeding materials.

Core Idea

Indirectly determine and select target traits by detecting **molecular markers (M)** tightly linked to the target (T), without directly observing the trait itself.

Key Premise

There must be a **tight and stable linkage** between the marker and the target gene to ensure accurate reflection.

Selection Advantage

Selection can be performed at the **early generation (like seedling stage)** without waiting for trait expression, significantly shortening the breeding cycle.

Bridging Genotype and Phenotype

Genotype
Invisible Genetic Essence

MAS Marker-assisted Selection
Precise Selection · Efficient Pyramiding

Phenotype
Visible Trait Expression

Breaking Environmental Barriers

Genotype-based selection avoids environmental noise, ensuring more accurate, stable, and reliable results.

Early Selection

Early seedling detection removes poor plants, cutting later field costs.

Rapid Gene Pyramiding

Multiplex background selection rapidly pyramids multiple target traits into a single plant.

Improving Breeding Efficiency

Shorter cycle, less guesswork, higher precision and efficiency.

Core Value: Turning “invisible” genotypes into “detectable” markers – breaking time and space limits for a revolutionary leap in breeding.

Development History of MAS Technology

1980s

Start of Molecular Markers

RFLP markers kicked off the molecular marker era and laid the technical foundation.

1990s

Popularization of PCR Technology

PCR technology drove the widespread use of PCR-based markers like AFLP and SSR.

2000s

Whole Genome Sequencing

Whole-genome sequencing of crops like rice gave us massive markers and genetic data.

2010s

NGS and Genomic Selection

NGS made SNP markers mainstream and sparked Genomic Selection (GS).

2020s

Precision and Smart Breeding

CRISPR and AI are now driving us into a new era of precision and intelligence.

Core of Technological Evolution: From single marker assistance to whole-genome information utilization, and further to precision design breeding driven by multi-omics and intelligent algorithms.

What can MAS do?

Gene Mapping

Construct genetic maps to precisely locate QTLs and target genes.

Germplasm Resource Identification

Evaluate genetic diversity and construct variety fingerprint maps.

Backcross Breeding

Foreground selection for target genes and background selection to recover the recipient genome.

Gene Pyramiding breeding

Pyramid multiple disease-resistant or high-quality genes into a single variety.

Early Generation Screening

Eliminate inferior plants at the seedling stage, significantly saving field planting costs.

Case 1: MAS for disease-resistant rice breeding

Background & Challenges

Bacterial blight: a widespread rice disease that can cause 20%–50% yield loss or crop failure.

Traditional breeding has a long cycle and is difficult to improve quickly.

Cross & Backcross

Resistant Donor × Elite Recipient

MAS Screening

Screening Xa21-containing plants using SSR markers in each generation.

Phenotypic Selection

Simultaneously selecting individuals with excellent agronomic traits.

Application Results

Successfully developed new rice varieties with both disease resistance and high yield.

It only takes **3-4** generations, shortening the time by more than **50%** compared to traditional breeding.

03. Techniques: Molecular Markers

Molecular Marker Technology and Genotyping

Classification of Molecular Markers

Hybridization-based markers: RFLP, detecting polymorphisms via molecular hybridization.

PCR-based markers: SSR, AFLP, ISSR, Indel, detecting polymorphisms via PCR amplification.

Sequencing-based markers: SNP, detecting single nucleotide polymorphisms via sequencing.

Genotyping

Genotyping involves detecting the genotype of each individual in a mapping population using molecular markers to obtain complete genotypic data, which is the basis for constructing genetic maps and QTL mapping.

Basic Characteristics of Ideal Molecular Markers

High Polymorphism

Detect extensive genetic variation to distinguish different genotypes.

Codominance

Clearly distinguish homozygotes and heterozygotes for complete information.

Good Stability

Unaffected by environmental factors with high experimental repeatability.

Uniform Distribution

Uniform coverage across the genome for overall scanning.

Easy Detection

Simple and rapid process, low cost suitable for large-scale applications.

No Phenotypic Effect

Does not affect target trait expression, only serves as a genetic marker.

Core Value: These characteristics collectively determine the reliability and efficiency of markers in genetic analysis and breeding applications, serving as key considerations when selecting molecular markers.

SSR: Simple Sequence Repeat

Definition

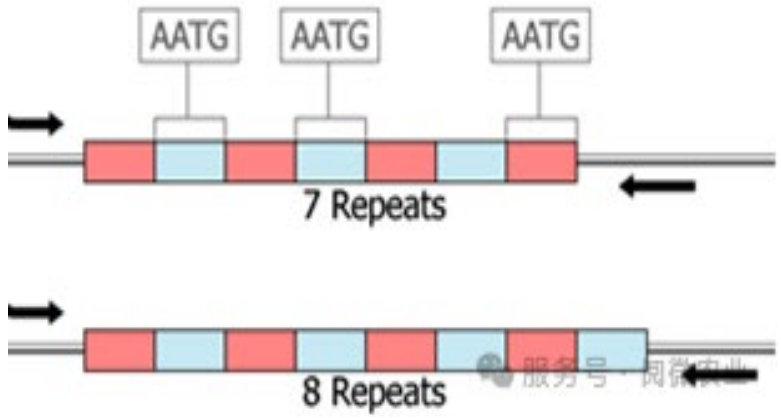
Tandem repeated sequences of 1-6 bases in the genome, also known as microsatellites.

Polymorphism Source

Differences in the number of core sequence repeats among individuals lead to varying amplified fragment lengths.

Technical Principle

Specific primers are designed based on flanking conserved sequences, and fragment length polymorphism is detected after PCR amplification.

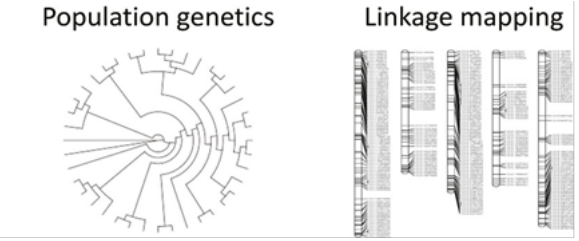


Core Value: SSR markers are codominant, highly polymorphic, and reproducible, widely used in genetic map construction, variety identification, and genetic diversity analysis.

SSR Technology

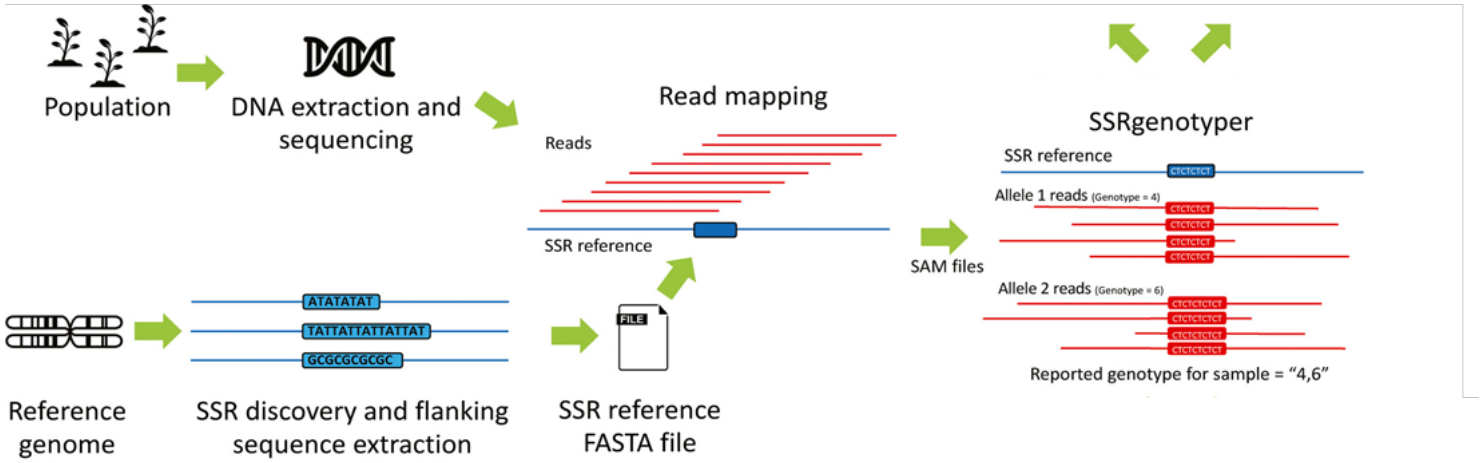
Advantages

- High polymorphism – abundant genetic info, high resolution
- Codominant – accurately distinguishes homozygotes from heterozygotes
- Good stability – mature tech, strong repeatability, reliable results
- Wide distribution – uniform across eukaryotic genomes



Limitations

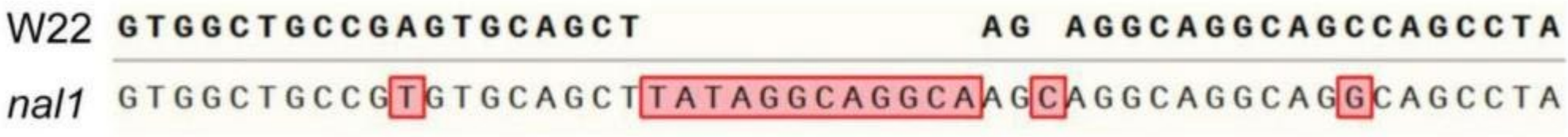
- High cost – needs flanking sequences; primer dev is slow and laborious
- Species-specific – low transferability; new primers typically required



SSR markers can be used to construct linkage maps and conduct population genetics analyses

InDel: Insertion Deletion Markers

- **Length polymorphism caused by the insertion or deletion of a DNA fragment at a specific position in the genome**



Key Characteristics

- **Wide distribution:** Abundant across the entire genome
- **High stability:** Low mutation rate, genetically stable
- **Flexible detection**: Compatible with both low-cost electrophoresis and high-throughput sequencing
- **Easy genotyping:** Most loci are biallelic, simplifying statistical analysis

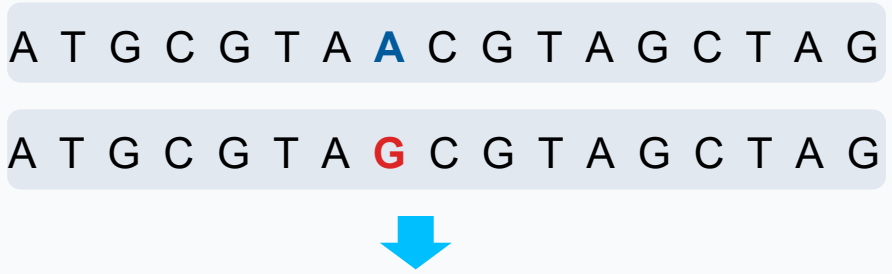
Main Application Areas

- Genetic map construction
- Marker-assisted selection (MAS)
- Germplasm identification
- Gene mapping

SNP: single nucleotide polymorphism

Definition: SNP refers to variations in a single nucleotide in the genome, including single-base substitutions, insertions, or deletions.

SNP Locus Principle Diagram



Abundant Quantity: Widely distributed across the genome, reaching tens of millions.

Widespread Distribution: Uniformly distributed in coding and non-coding regions.

High-throughput Detection: Compatible with gene chips and sequencing, high automation.

Genetic Stability: Low mutation rate, stable and reliable genetic information.

Entering the Genome-wide Era

The high-density nature of SNPs enables Genome-wide Association Studies (GWAS).

Limitations

- Each marker provides relatively little information
- Requires high-density coverage for effective selection

SNP Discovery and Validation Process

01 SNP Discovery Strategies

Re-sequencing

By deep sequencing whole genomes or targeted regions across multiple individuals and directly comparing sequence variations, SNPs can be discovered.

Reduced-Representation Sequencing

GBS or RAD-seq reduces complexity by enzyme digestion, allowing efficient, low-cost discovery of large numbers of SNPs.

Genotyping by Array

Using high-density SNP arrays for population genotyping quickly uncovers known and new variant loci.

02 SNP Validation and Evaluation

Validation Purpose: Validate candidate SNP loci in larger natural or mapping populations to **exclude sequencing/analytical errors**, ensuring genuine loci with stable polymorphism. Common methods include Sanger sequencing (gold standard), KASP, CAPS, etc.

Discovery (from massive data) → Validation (experimental, for reliability) → Application (genetic analysis & breeding)

SNP Detection Technology (I): High-Throughput

Gene Chip Technology (SNP Chip)

Principle: Immobilize a large number of **SNP probes on a solid-phase carrier, hybridize with fluorescently labeled sample DNA**, and determine genotypes based on signal intensity.

Features: Extremely **high throughput**, capable of detecting thousands of loci at once, suitable for large-scale population genotyping.

Applications: Genome-wide association studies (GWAS), genomic selection (GS).

Next-Generation Sequencing (NGS)

Principle: Directly read DNA sequences **through large-scale parallel sequencing** to discover and identify SNP loci.

Features: **Highest resolution**, capable of discovering new SNPs, flexible detection but relatively **high cost**.

Applications: Whole-genome resequencing, genotyping by sequencing (GBS/RAD-seq), targeted sequencing.

Technology Comparison Summary: Chip technology focuses on "cost-effective large-scale screening" for known loci; sequencing technology emphasizes "high-resolution full-sequence exploration" for discovering new variants.

SNP Detection Technology (II): Medium-Low Throug

KASP Technology

A PCR-based technology that uses primer competition: specific primers are designed for the two SNP alleles, and genotypes are determined by fluorescent signals.

Features: High accuracy, high flexibility, moderate cost, suitable for medium-low throughput genotyping.

CAPS / dCAPS

Uses SNP to alter restriction enzyme recognition sites, detected by PCR and electrophoresis. dCAPS creates sites via mismatches.

Features: Low cost, simple operation, suitable for routine lab testing.

SSCP Technology

Detects SNPs based on altered electrophoretic mobility due to single-stranded DNA conformational differences.

Features: Simple and fast, but relatively low resolution.

Application Summary: For a small number of loci, CAPS or SSCP is preferred to control costs; for medium-throughput loci requiring high accuracy, KASP is a better choice.

Application of SNP

Construction of High-Density Genetic Maps

High-density genetic linkage maps have been constructed for various crops such as apple, jujube, and longan using SNP markers, laying a solid foundation for gene mapping.

Genome-Wide Association Study (GWAS)

In crops like rice, maize, and wheat, SNP chips have been successfully used to identify numerous key genetic loci controlling complex traits such as yield, quality, and resistance.

Genomic Selection (GS)

In plant breeding, prediction models are built using genome-wide SNP markers to enable early and efficient selection of complex traits, significantly shortening the breeding cycle.

Marker-Assisted Selection (MAS)

SNP markers closely linked to target genes are converted into KASP or CAPS markers for precise selection of breeding materials, improving breeding efficiency.

Application Cases of Molecular Marker

Case	Species	Target Nutritional Trait	Marker Type	Key Conclusion	Reference
1	Wheat	Fe, Zn	SSR	SSR markers significantly associated with Fe/Zn content; can be used in MAS for high-micronutrient genotypes	Heidari B. et al., *Sci Rep*, 2024
2	Winter squash	Starch, carotenoid	InDel	253 InDel markers mapped 22 QTLs; PVE up to 15.13%	*Duan Y.* et al., *Euphytica*, 2025
3	Rice	Low phytate (Fe/Zn bioavailability)	CAPS	CAPS marker 100% linked to the *OsMRP5* mutation site	Zia-ul-Qamar et al., *Front Plant Sci*, 2024
4	Soybean	Protein	InDel + dCAPS	InDel (low band = high protein) + dCAPS combination; 533 polymorphic markers; 6 QTLs detected	Wang et al., *Sci Agric Sin*, 2019
5	Wheat	Protein quality (gluten)	KASP (11)	97 SNPs→43 QTLs; 11 KASP markers accuracy >85%	Zhang et al., *BMC Plant Biol*, 2025,

Case GWAS & Genomic Selection for Southern Corn Rust Resistance

Context: SCR causes 20–50% yield loss. Two major resistance genes *RppC* and *RppK* (NLR-type) are known.

Materials: 384 DH lines + 903 testcross hybrids

Phenotypic data Collection and analysis

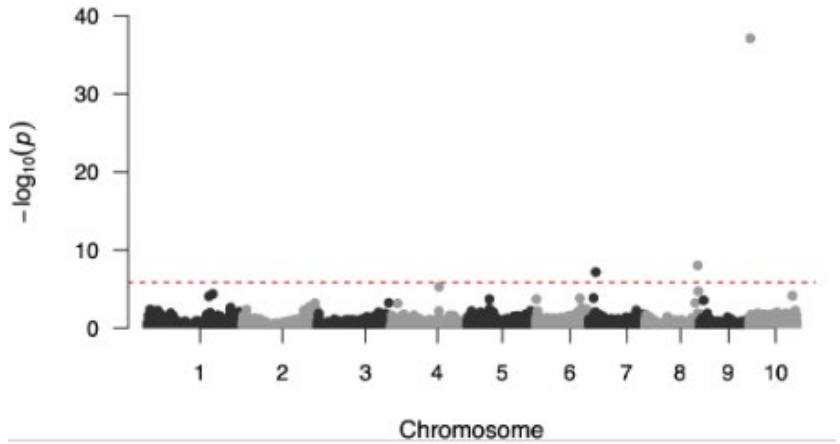
the SCRRSs in DH founders and testers.

Genotyping and genotypic data analysis

Genotyping used the **Maize-6H-60K SNP chip**.

GWAS and Genomic prediction

- Chr10 QTL tightly linked to RppC / RppK
- Genomic prediction used three models.



Model /	Prediction accuracy /
GBLUP	0.56 – 0.60
BayesA / BayesB	0.65
Bayes + QTL as fixed effect / Bayes + QTL	**0.67** (↑11.7%)

Li Y, et al. *Frontiers in Plant Science*, 2023

Core: GWAS captures major R genes. Including QTL information significantly improves GS accuracy

04. Validation: Candidate Gene Validation

Candidate Gene Discovery (I): Forward Genetics

Method	Principle	Application	Resolution
Map-based cloning	Fine mapping using molecular markers	Model plants with high-quality genomes	Single gene
BSA (Bulked Segregant Analysis)	Extreme-phenotype pools for rapid interval mapping	Major genes, preliminary localization	Mb-level
MutMap	Whole-genome resequencing of mutant pools	Mutants with reference genome	Single base
QTL mapping	Linkage analysis in segregating populations	Quantitative traits (yield, disease resistance)	cM- or Mb-level
GWAS **BSR-seq / QTL-seq**	LD-based association in natural populations BSA + RNA-seq or whole-genome resequencing	Germplasm panels Non-model species or with expression data	Single gene or fine interval Candidate gene/interval

Candidate Gene Discovery (Ⅱ): Reverse Genetics

Method	Principle	Application	Resolution
CRISPR/Cas9	Targeted genome editing via double-strand breaks	Agrobacterium-mediated transformation, protoplasts	Efficient, precise
RNAi	Gene silencing via dsRNA	Stable transgenic lines	Inducible, tissue-specific
VIGS	Transient gene silencing using viral vectors	TRV vector, Agrobacterium infiltration	Rapid, no stable transformation
TILLING	Point mutation screening in mutagenized populations	EMS population + high-throughput sequencing	Non-transgenic, allelic series
Overexpression	Constitutive expression under strong promoter	Overexpression vector, Agrobacterium transforma tion	Gain-of-function validation
Activation tagging	Random T-DNA insertion with enhancers	Transgenic screening	Identifies redund

Functional Verification Methods for Candidate

Transgenic Technology

Principle: Introduce candidate gene coding sequences into recipient plants via Agrobacterium-mediated transformation or gene gun to achieve overexpression or silencing, then observe phenotypic changes.

Characteristics: Mature technology with wide applications; however, it has a long cycle and may have position effects.

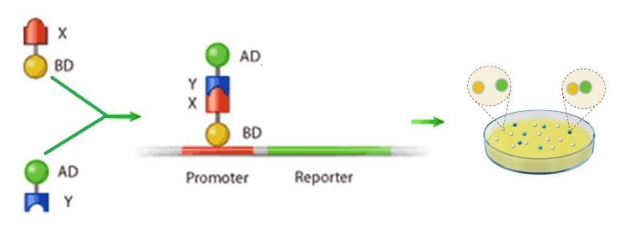
Gene Editing (CRISPR/Cas9)

Principle: Use the CRISPR/Cas9 system for site-directed mutagenesis, knockout, or insertion of candidate genes, and determine gene function through mutant phenotype analysis.

Characteristics: Precise, efficient, short cycle, mimics natural mutations, and is the most advanced verification tool currently.

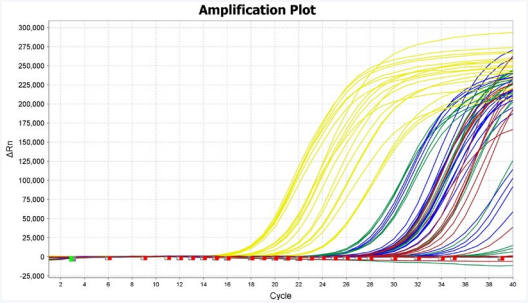
Core Value: Tech iteration from overexpression to precision mutation – faster, more accurate functional validation, powering modern molecular biology.

Functional Verification Methods for Candidate Gene



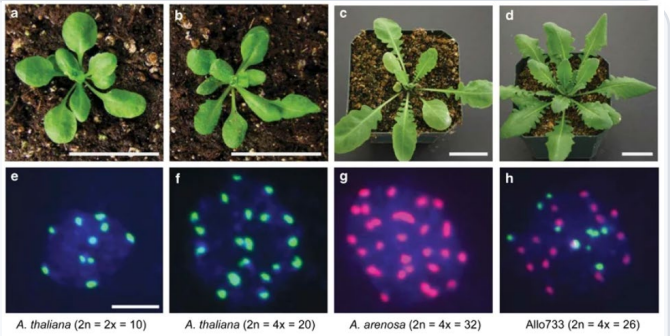
Yeast Two-Hybrid

Principle: Verify protein-protein interactions to reveal molecular mechanisms and pathways of gene regulation.
Features: A classic method for studying protein-protein interactions.



Gene Expression Analysis

Principle: Detect expression levels under different spatiotemporal conditions using techniques like qRT-PCR.
Features: Rapid and simple, serving as the foundation for gene function research.



In Situ Hybridization

Principle: Use labeled probes to hybridize with tissue sections and determine the specific location of gene expression.
Features: Intuitively display the spatiotemporal expression pattern of genes.

Complete Process of Candidate Gene Validation

Phenotypic and Genetic Analysis

Clarify the genetic basis of target traits, distinguishing between qualitative and quantitative traits.

Gene Mapping

Use QTL mapping or GWAS to localize the target trait to a specific chromosomal region.

Candidate Gene Discovery

Screen potential candidates based on genome annotation, transcriptomics, and homology.

Functional Validation

Validate gene function using overexpression, CRISPR knockout, and expression analysis.

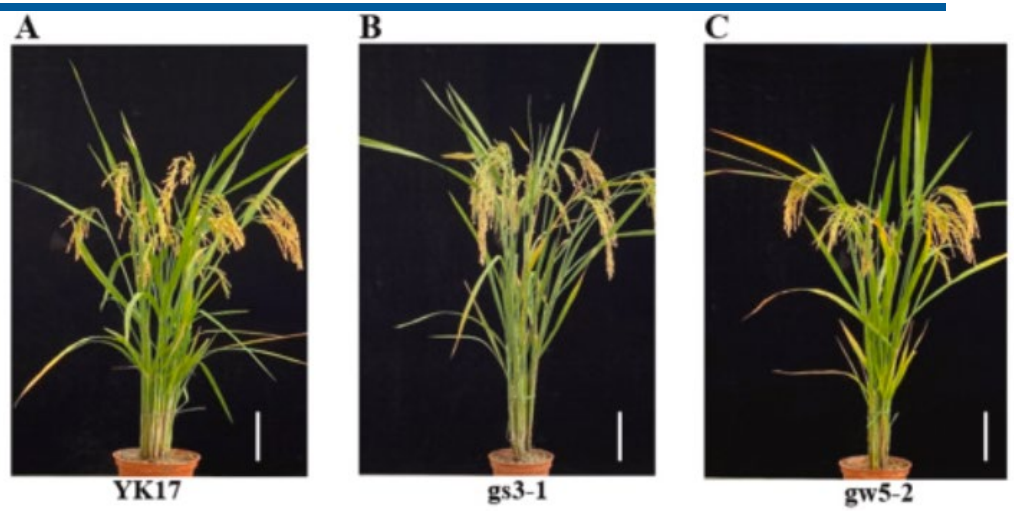
Marker Development and Application: Convert validated genes into efficient markers for Marker-Assisted Selection (MAS) to accelerate breeding.

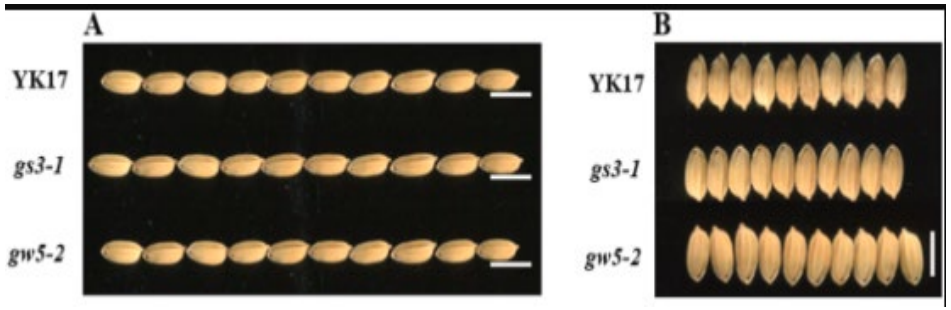
Multiplex CRISPR Editing of GS3 and GW5 in Rice

Target genes: GS3 – negative regulator of grain length,
GW5 – negative regulator of grain width & weight

Strategy:

- Two sgRNAs targeting first exons → frameshift → null alleles
- Single vector, Agrobacterium-mediated transformation





Key insight:

- Both genes are negative regulators.

Phenotypes of single knockouts					
Mutant	Grain length	1000-grain weight	Chalkiness	Amylose	Panicles per plant
gs3-1	↑↑↑	–	↓↓	↓	↓
gw5-2	–	↑↑	–	↑	↓

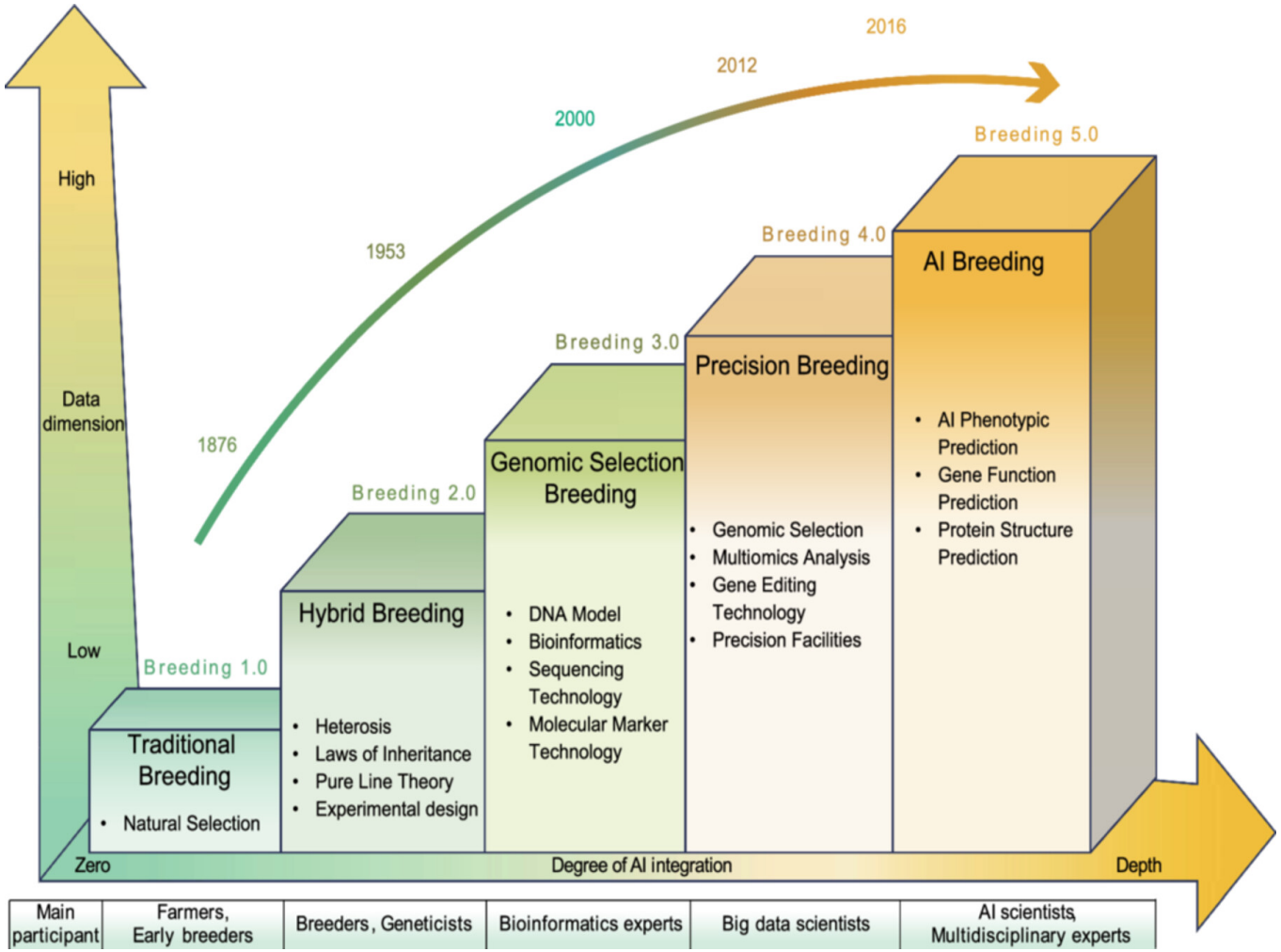
- Grain size improvement can trade off with panicle number → need optimal background.

Breeding value: One-step generation of elite alleles; faster than conventional pyramiding

Aqib Zeb, et al. *Ecological Genetics and Genomics*, 2025

05. Systems: Decision Support Systems

Evolution of breeding technology: From empirical practices to AI-driven paradigms



Breeding 1.0 and 2.0 mainly relied on empirical phenotypic analysis, but subsequent stages—including genomic selection, marker-assisted breeding, multi-omics data integration, and deep learning prediction models—have continuously improved efficiency and accuracy through data-driven technological innovation.

Decision Support System: Concept and Architecture

A system that integrates multi-dimensional data (genotype, phenotype, environment) using computer technology, big data analytics, and AI algorithms to deliver intelligent decision support for breeders.

Data Layer

Integrates germplasm resources, genotype, phenotype, and environmental meteorological data for unified management.

Model Layer

Constructs genomic selection, phenotype prediction, and parent selection models to provide scientific prediction capabilities.

Application Layer

Provides a user-friendly visualization interface for data query, analysis, and intelligent breeding scheme recommendation.

Key Functional Modules of the Decision Support System

Germplasm Resource Management

Digitally manage and query basic information, phenotypic data, and genotypic data of germplasm resources.

Parent Selection Recommendation

Recommend optimal parental combinations based on breeding objectives and genetic distance.

Progeny Performance Prediction

Predict phenotypic performance of hybrid progeny using genomic selection models to screen potential individuals.

Field Trial Design

Optimize field trial layout, replication, and management to improve efficiency.

Data Analysis & Visualization

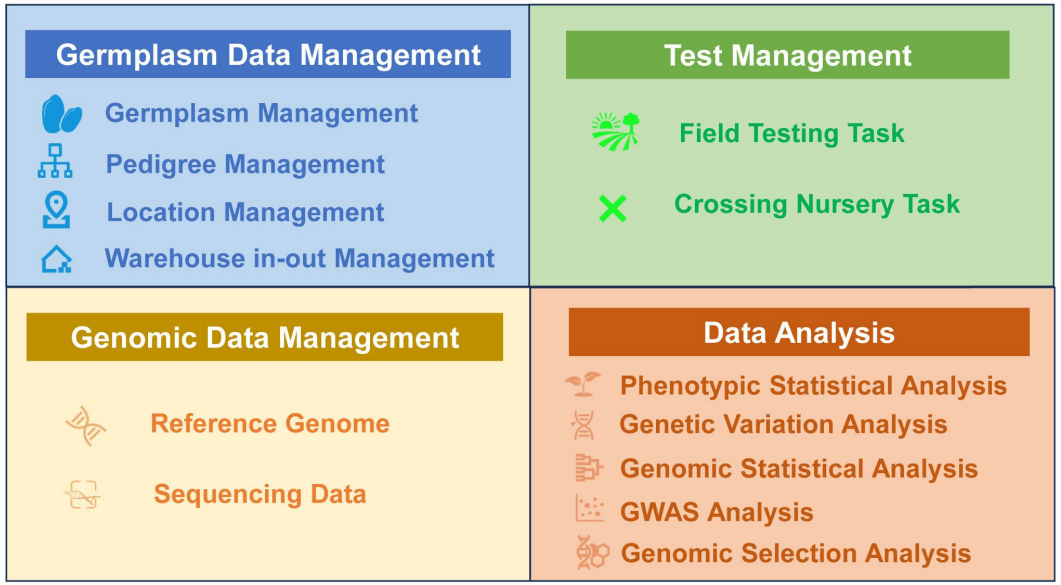
Perform statistical analysis and visualization of breeding data to identify patterns and trends.

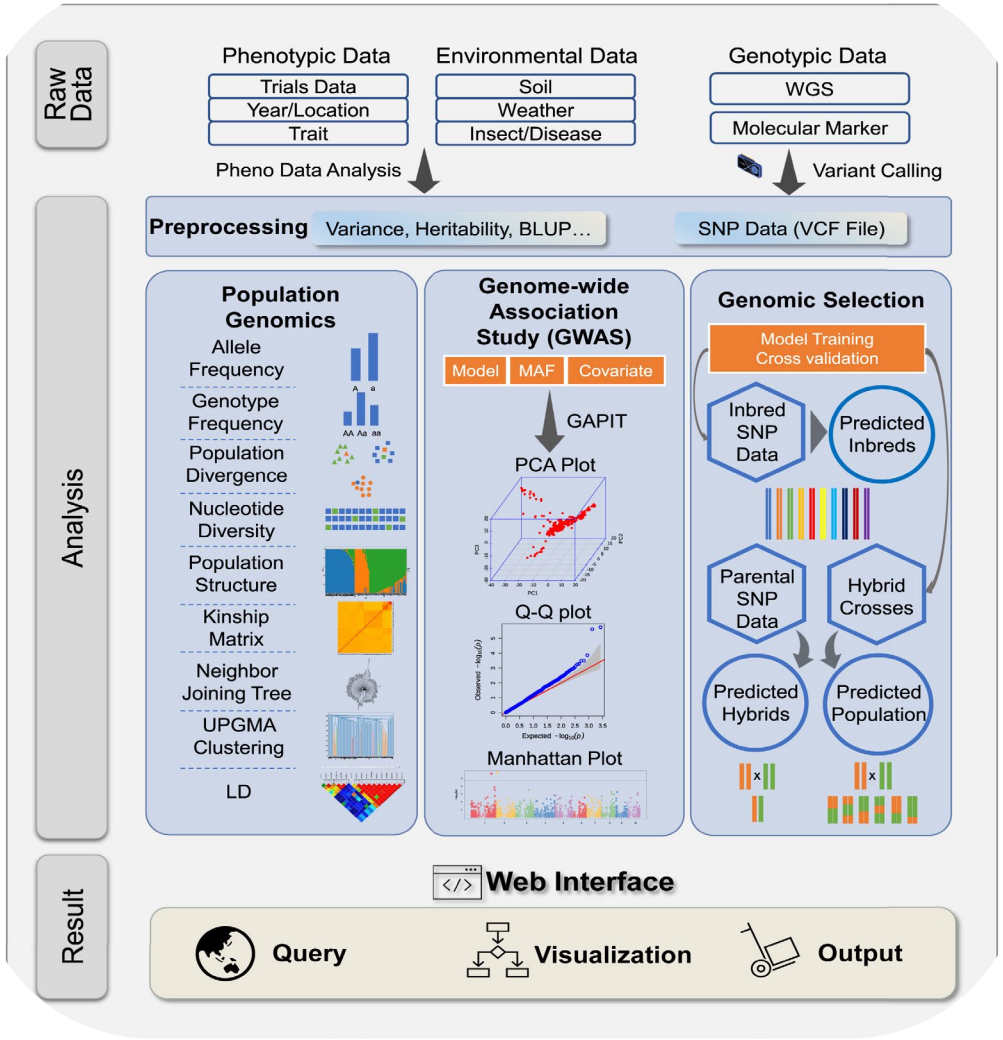
Intelligent Decision Support

Integrate multi-dimensional data to assist breeders in making more scientific and efficient decisions.

Case Study: Smart Breeding Platform

the National Nanfan Research Institute (CAAS) and Alibaba, among others, have launched the **Smart Breeding Platform** (https://sbp.ibreed.cn) – a free, web-based tool for managing and analyzing large-scale genetic, genomic, and phenomic data.





06. Outlook: Challenges & Future Directions

Current Challenges

Complex Trait Analysis

Minor-effect genes create complex networks that hinder deciphering key agronomic traits.

Gene-Environment Interaction

Environment strongly affects gene expression, making G×E prediction a major challenge for stable breeding

Technology Cost & Accessibility

Sequencing and gene editing are costly, restricting use in poorer regions and minor crops.

Biosafety & Regulation

Ethical concerns and biosafety gaps remain for new technologies.

Data Integration & Sharing

Lack of standards hinders integration of heterogeneous data and cross-platform analysis.

Conclusion: To achieve sustainable molecular breeding, we need stronger basic research, lower costs, better regulations, and data sharing.

Future Development Trends

Popularization of Genomic Selection

Routine, early, efficient, precise selection in breeding.

Precision of Gene Editing

CRISPR: more precise, efficient, and widely used for crop improvement.

Deep Integration of Multi-omics

Integrating genomic, transcriptomic, and other multi-dimensional data to comprehensively analyze life processes.

Deep Application of AI

Drives gene discovery, breeding decisions, and intelligent development.

Intelligent Design Breeding

From empirical breeding to precision design for targeted creation of new varieties.

Summary: Technology integration and intelligence are the core driving forces of future breeding, which will realize the leapfrog development from "experience" to "precision".

Introduction to Genomic Selection (GS)

Core Principles

Using genome-wide SNP markers and training population data, prediction models estimate GEBVs for early, efficient selection.

Technical Advantages

Uses genome-wide info to shorten cycles, boost accuracy. Ideal for complex, low-heritability, hard-to-measure traits.

Differences from MAS

MAS

Only targets a few major QTLs/genes; limited effect on complex traits.

GS

Uses all genome-wide markers, including minor QTLs, to effectively improve multi-gene complex traits.

Summary: GS is a breakthrough that uses all genome-wide markers to overcome traditional MAS limitations in improving complex traits.

Gene Editing Technology in Breeding

Core Principles of CRISPR/Cas9

gRNA guides Cas9 to cut specific DNA sites. Repair enables knockout, insertion, or replacement for targeted editing.

Core Applications in Breeding

Rapid improvement – knock out bad genes for faster trait gains.
Precise transfer – insert or swap good genes for directed evolution.
Safety – no foreign genes, better acceptance.

Latest Technological Breakthroughs

Base editing – single-base swap, no double-strand breaks, fewer off-targets.
Prime editing – any base, insertion, or deletion, higher precision.

Core Competitive Advantages

Gene editing is more precise, efficient, and cost-effective than traditional breeding.

Conclusion: Gene editing is integrating with technologies like Marker-Assisted Selection (MAS) to usher in a new era of precision crop breeding.

Multi-omics Technology and Breeding

Genomics

Analyze whole-genome DNA sequences to discover all genes and genetic markers, serving as the foundation for all omics studies.

Metabolomics

Qualitatively and quantitatively analyze all metabolites in organisms to establish direct links between genotypes and phenotypes.

Transcriptomics

Profile all RNA transcripts in specific tissues or developmental stages to reveal gene expression patterns and regulatory networks.

Multi-omics Application

Integrate genomics, transcriptomics, proteomics, and metabolomics data to comprehensively dissect complex trait formation mechanisms.

Proteomics

Study expression levels and post-translational modifications of all proteins in cells to understand the direct executors of life activities.

Accelerate Gene Discovery and Molecular Breeding

Application of Machine Learning in Breeding

Complex Phenotype Prediction

Construct models using multi-omics data to accurately predict complex phenotypes such as yield and quality, significantly reducing the time and economic costs of field identification.

Gene-Phenotype Association Analysis

Mining non-linear relationships in data through deep learning to help scientists quickly locate gene loci controlling key traits.

Intelligent Breeding Decision Making

Optimize parent selection schemes, accurately predict hybrid offspring performance, and assist in formulating more scientific and efficient breeding strategies.

Automated Phenotyping Analysis

Combining computer vision technology to achieve automatic identification and accurate measurement of crop phenotypes such as plant height, leaf area, and pests/diseases.

Data-Driven Breeding to Intelligent Design

Comprehensive Data Collection System

- **Satellite Remote Sensing** – Macroscopic monitoring of large-scale crop growth environments and status.
- **UAV Networking** – Mesoscale acquisition of field crop phenotypes and pest data.
- **Multi-omics Technology** – Microscopic analysis of molecular mechanisms underlying genotype-environment interaction.
- **Massive Database Construction** – Integrating genotype, phenotype, and environmental data to lay the foundation for intelligent breeding.

Intelligent Design and Prediction

- **Deep Mining and Simulation** – Virtual environments for simulating gene combination performance under complex conditions.
- **Accurate Prediction of Breeding Outcomes** – Eliminates >90% of ineffective plans upfront, greatly boosting screening efficiency.
- **Self-Reinforcing Closed Loop** – “Collect → Model → Verify → Iterate” for continuous model optimization.
- **From Experience to Intelligence** – Shifting breeding from field screening to intelligent design.

International Development Trends in Molecular Bre

Giants Accelerate Technology Layout

Seed industry giants such as Bayer and Corteva have deeply integrated gene chip technology into their core breeding processes, driving industry-wide technological transformation.

Construction of Precision Breeding Platforms

Building "genotype-phenotype" big data platforms, such as Bayer's Preceon system, significantly improves the efficiency and predictability of genetic improvement.

Accelerated Commercialization

India has launched the world's first gene-edited rice varieties, and the US has approved gene-edited PRRS-resistant pigs for market, marking breakthroughs in regulation and application.

A New Phase of Global Competition

Molecular breeding technology is moving from laboratories to large-scale industrial applications, ushering in a new technology-driven phase in the global seed industry competition.

07. Summary & Acknowledgements

Summary

Marker-Assisted Selection (MAS)

Bridging genotype and phenotype, enabling precise and efficient breeding selection, it is a core technology of modern breeding.

Candidate Gene Validation

A key step from markers to genes and from association to function, providing precise targets for MAS.

Decision Support System

Integrating multi-dimensional data and utilizing artificial intelligence to achieve intelligent breeding decisions, it is the development direction of future breeding.

The combination of the three promotes crop breeding from traditional experience to precise design and intelligent creation, ensuring food security and sustainable development.

Thank You