USE OF GENOMICS IN THE IMPROVEMENT OF CROPS

Zenaida V. Magbanua
DOST and University of the Philippines
Balik-Scientist Awardee

PROBLEMS FACING THE 21ST CENTURY

- New diseases/pathogens
- Old diseases
- Climate change/global warming
- Astronomic population growth
  - Dwindling world resources

PROBLEMS FACING THE 21ST CENTURY

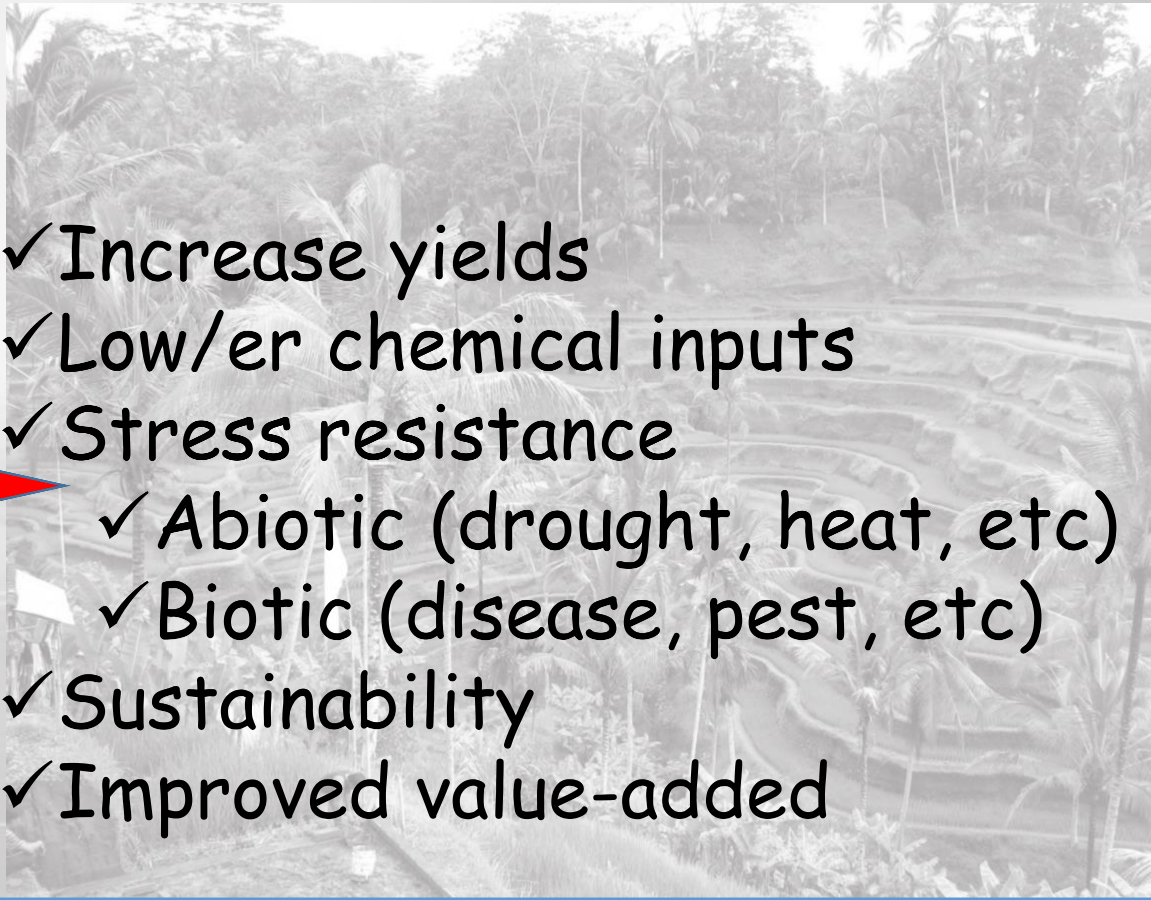
- New diseases/pathogens
- Old diseases
- Climate change/global warming
- Astronomic population growth
  - Dwindling world resources

agriculture

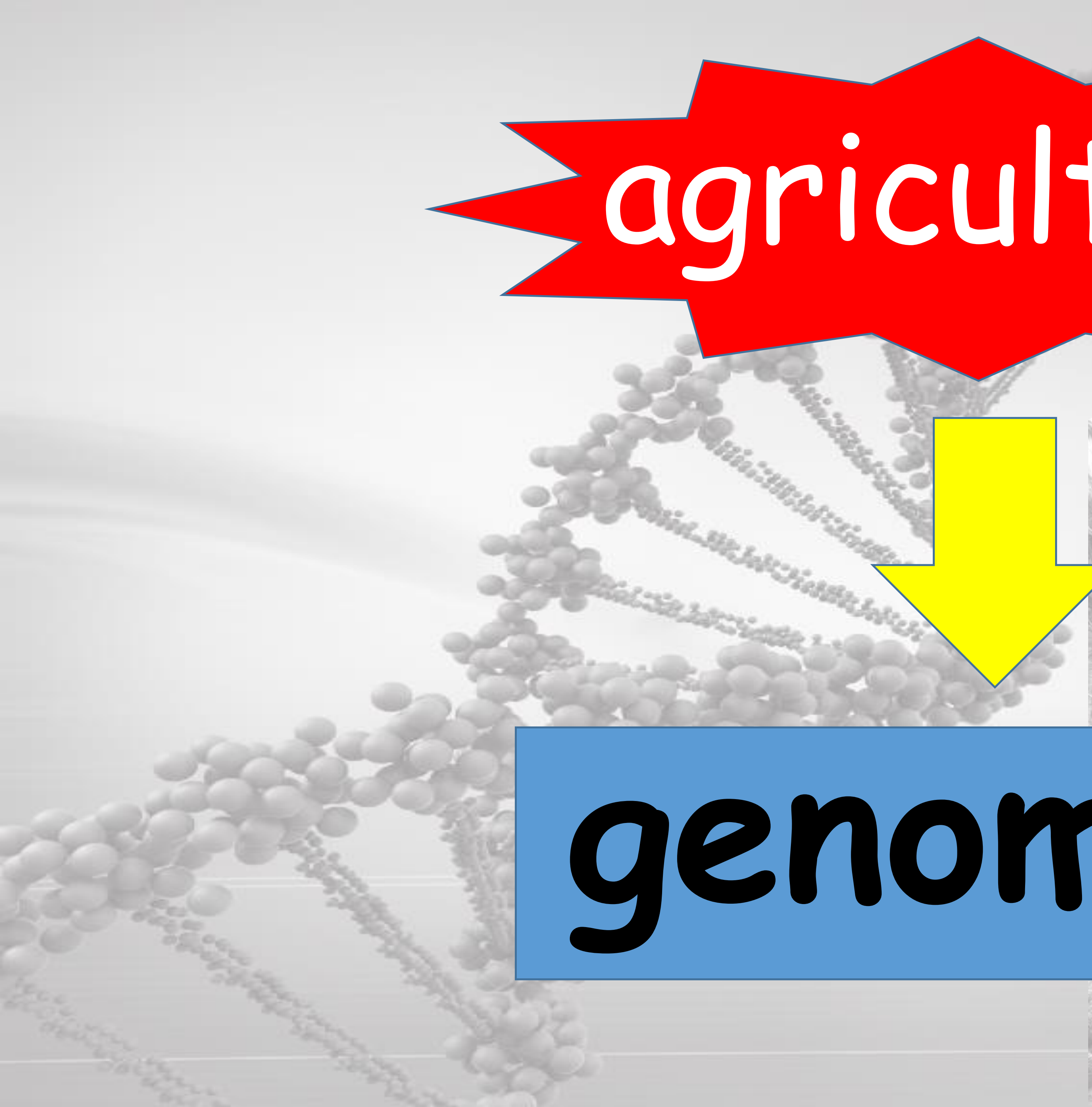
PROBLEMS FACING THE 21ST CENTURY

- New diseases/pathogens
- Old diseases
- Climate change/global warming
- Astronomic population growth
  - Dwindling world resources

agriculture



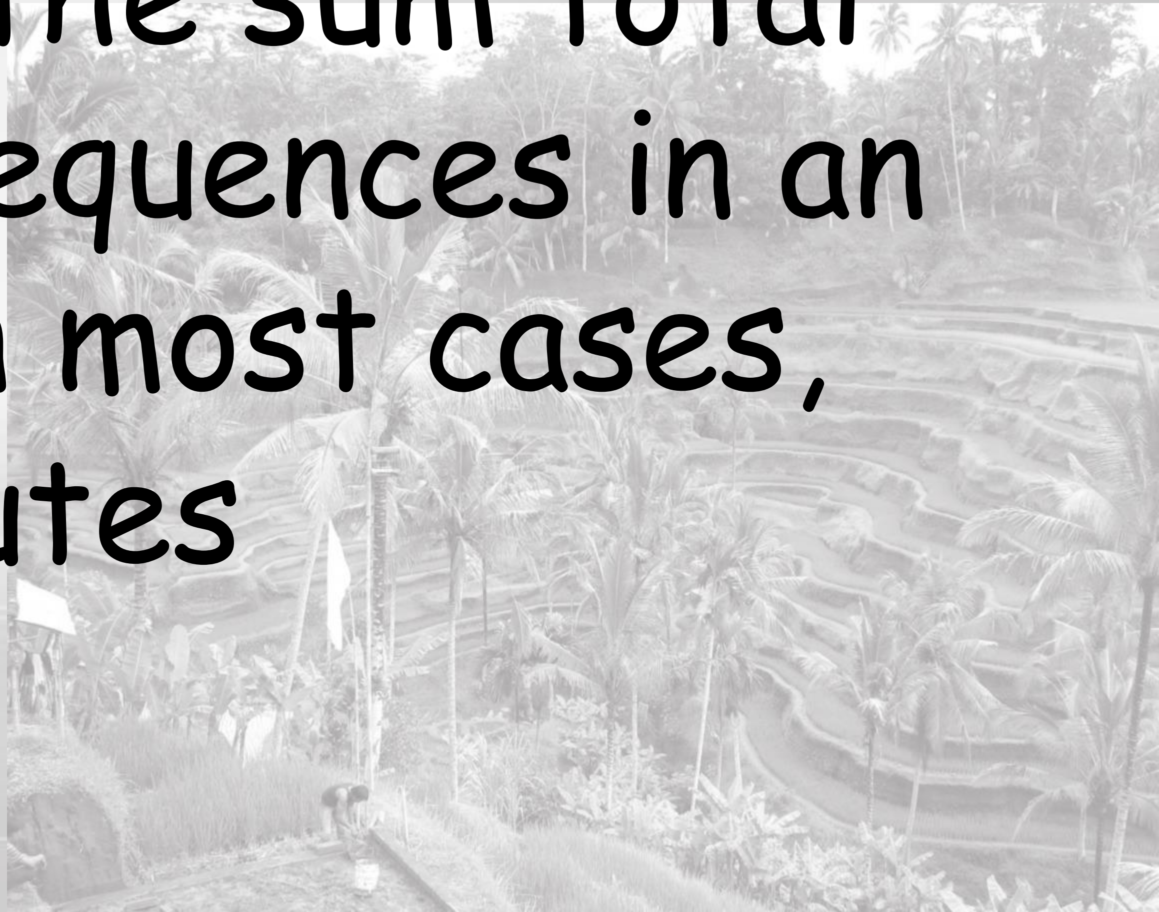
- ✓Increase yields
- ✓Low/er chemical inputs
- ✓Stress resistance
  - ✓Abiotic (drought, heat, etc)
  - ✓Biotic (disease, pest, etc)
- ✓Sustainability
- ✓Improved value-added



agriculture

genomics

Genomics: the study of the genome which is the sum total of all heritable sequences in an organism, DNA in most cases, and all its attributes

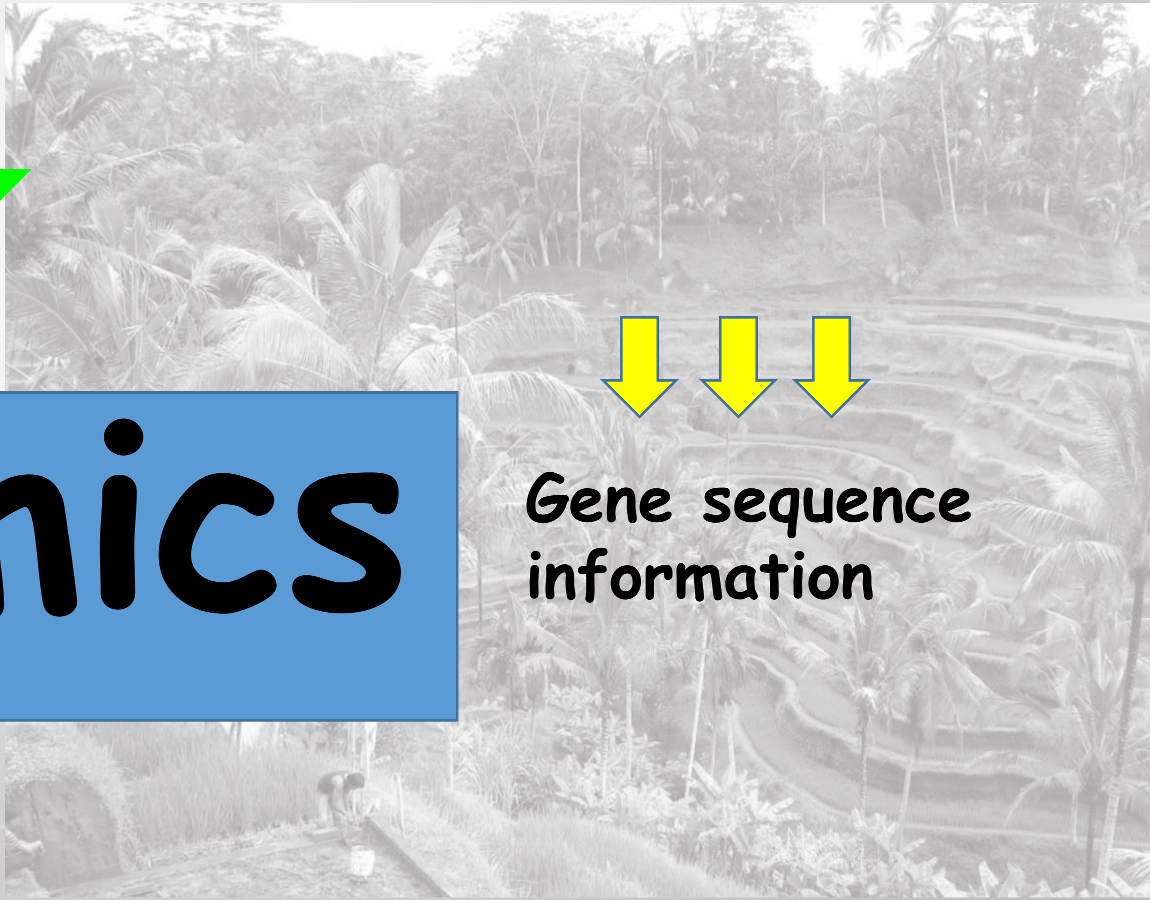


Traits that are important in agriculture are dictated by genes annotated from DNA sequences

- Increase yields
- Low/er chemical inputs
- Abiotic stress resistance
- Biotic stress resistance
- Sustainability

agriculture

genomics



Gene sequence information

Current research in Agriculture

Improvement of crop production through

- traditional breeding

Current research in Agriculture

Improvement of crop production through

- traditional breeding ➡ **SLOW**

Current research in Agriculture

Improvement of crop production through

- traditional breeding
- marker-assisted selection

Current research in Agriculture

Improvement of crop production through

- traditional breeding ➡ SLOW
- marker-assisted selection ➡ **BETTER BUT DOES NOT WORK ALL THE TIME**

Current research in Agriculture

Improvement of crop production through

- traditional breeding
- marker-assisted selection
- transgenic technology

Current research in Agriculture

Improvement of crop production through

- traditional breeding ➡ **SLOW**
- marker-assisted selection ➡ **BETTER BUT DOES NOT WORK ALL THE TIME**
- transgenic technology ➡ **LIMITED KNOWLEDGE ON GENE SEQUENCES**

Current research in Agriculture

Improvement of crop production through

- traditional breeding
- marker-assisted selection
- transgenic technology
- RNAi technology

Current research in Agriculture

Improvement of crop production through

- traditional breeding ➡ **SLOW**
- marker-assisted selection ➡ **BETTER BUT DOES NOT WORK ALL THE TIME**
- transgenic technology ➡ **LIMITED KNOWLEDGE ON GENE SEQUENCES**
- RNAi technology ➡ **LIMITED KNOWLEDGE ON GENE SEQUENCES**

Current research in Agriculture

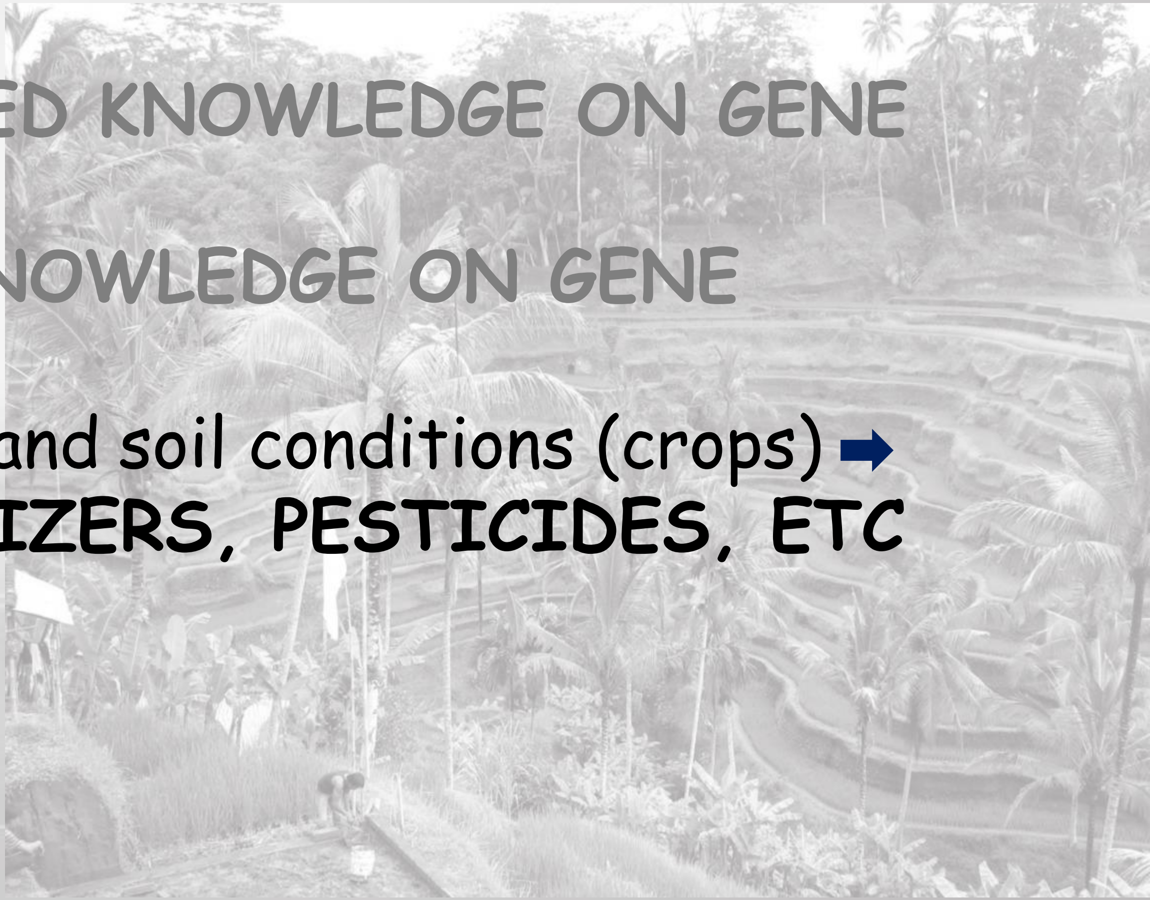
Improvement of crop production through

- traditional breeding
- marker-assisted selection
- transgenic technology
- RNAi technology
- optimization of chemical inputs and soil conditions

Current research in Agriculture

Improvement of crop production through

- traditional breeding ➡ **SLOW**
- marker-assisted selection ➡ **BETTER BUT DOES NOT WORK ALL THE TIME**
- transgenic technology ➡ **LIMITED KNOWLEDGE ON GENE SEQUENCES**
- RNAi technology ➡ **LIMITED KNOWLEDGE ON GENE SEQUENCES**
- optimization of chemical inputs and soil conditions (crops) ➡ **RELIES HEAVILY ON FERTILIZERS, PESTICIDES, ETC**



Traditional breeding

- Selects for desired trait/s by crossing and waiting for the progeny to grow and manifest phenotype

Traditional breeding for high yield and disease resistance

Disease resistant, but low yielding

Disease susceptible, but high yielding

Several generations, ie years of crosses

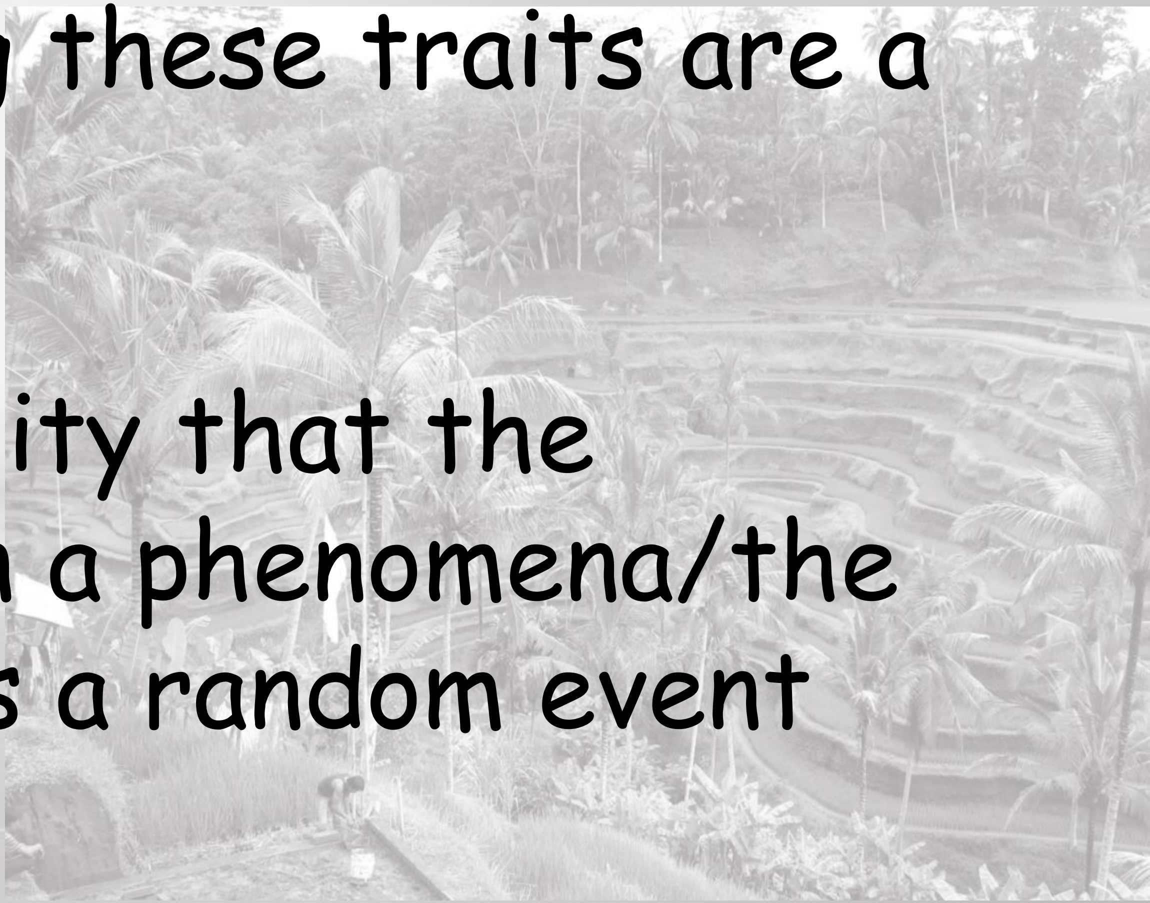
Disease resistant and high yielding

Marker-assisted selection

- Uses molecular markers to breed for desirable traits
  - Markers may be SSRs, RFLPs, AFLPs, etc
  - Usually applies to quantitative trait loci or quantitative traits (QTLs)

QTLs

- Traits whose expression are not determined by a single gene or locus
- Mapping and breeding these traits are a big challenge
- LOD – log of odds
  - log of the probability that the evidence can explain a phenomena/the probability that it is a random event



Examples of LOD

a) 1/1

LOD = 0

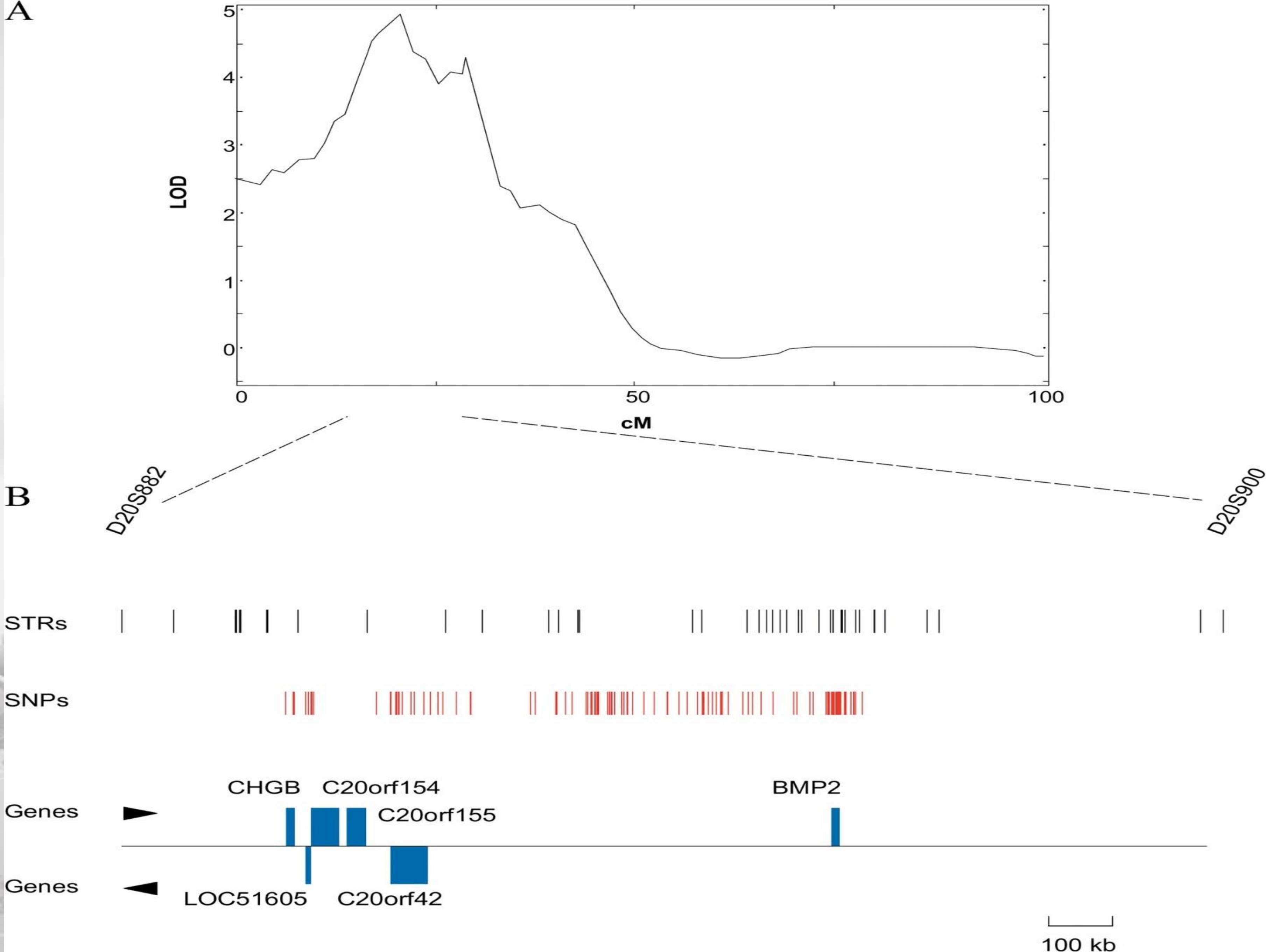
b) 1/0.1

LOD = 1

c) 1/0.01

LOD = 2

A cut-off of LOD 3 is usually accepted as support for linkage in genetics


A
LOD
5
4
3
2
1
0
0
50
100
cM
B
D20S882
D20S900
STRs
SNPs
Genes
Genes
CHGB
C20orf154
C20orf155
BMP2
LOC51605
C20orf42
100 kb

Examples of quantitative traits

- maize resistance to *Aspergillus flavus*

Examples of quantitative traits

- maize resistance to *Aspergillus flavus*
- rice resistance to bacterial panicle blight

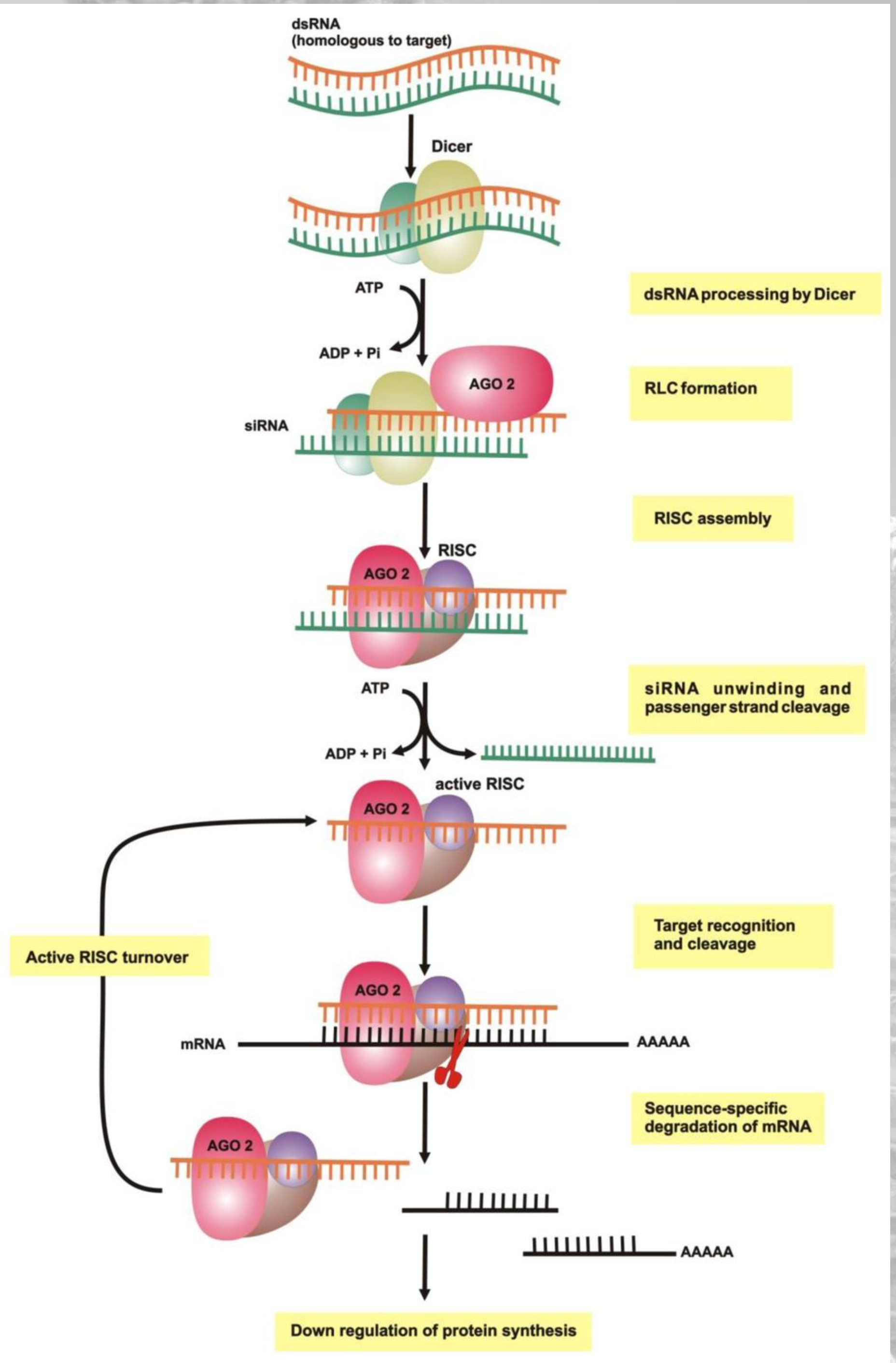
Transgenic technology

- aka genetic engineering technology
- transfer of a gene or a part thereof into the same or another organism

- <u>but we need to know the gene sequence</u>

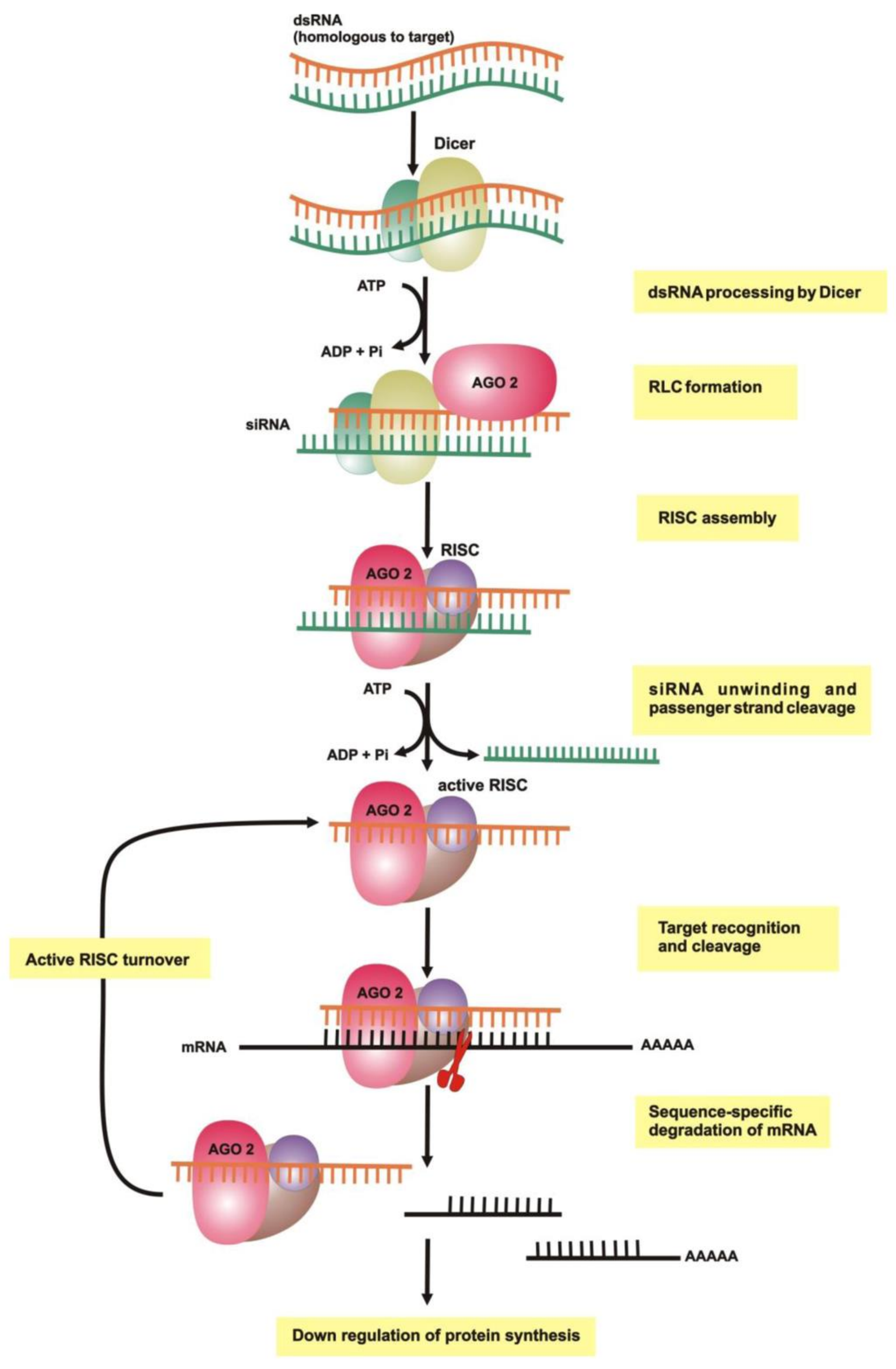
RNA interference

- aka gene silencing
- a mechanism acquired by genomes to ward off virus infection and transposons

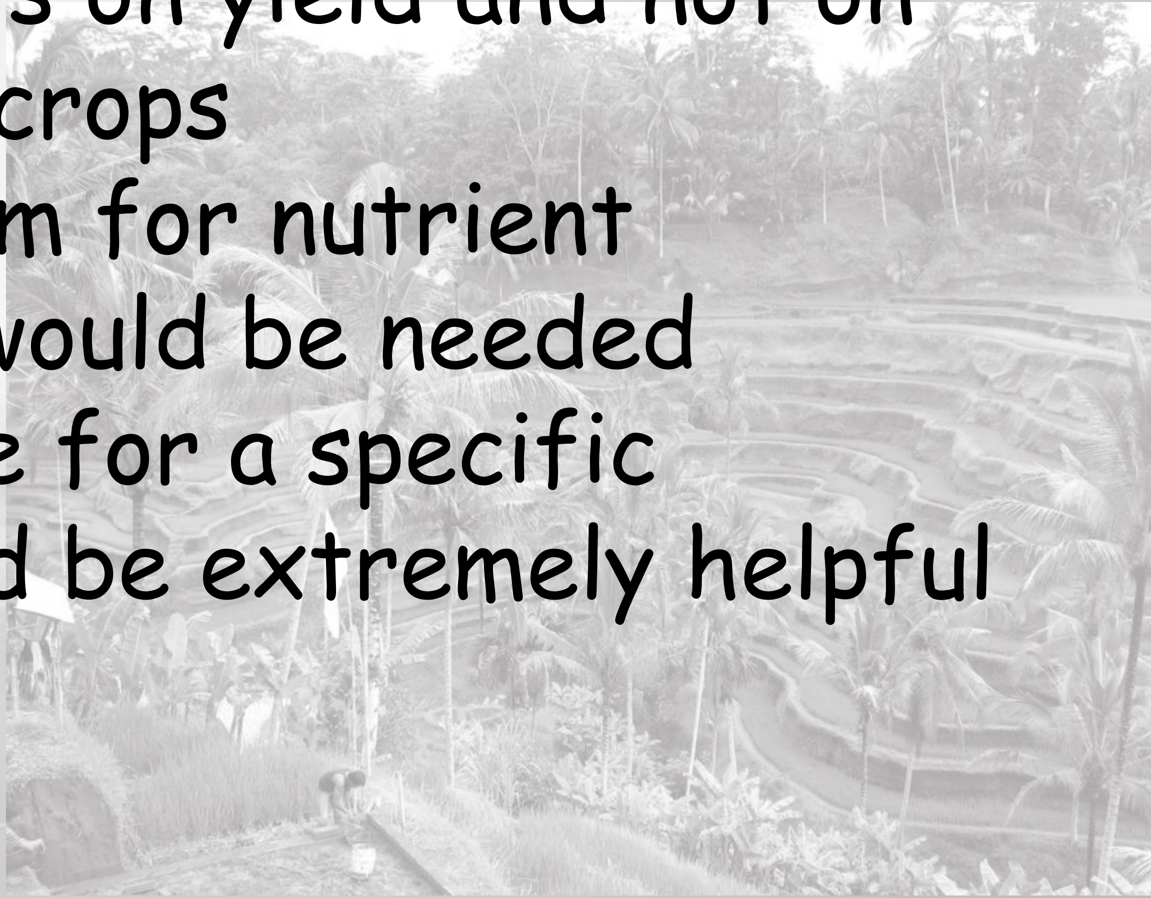


RNA interference

- aka gene silencing
- a mechanism acquired by genomes to ward off virus infection and transposons
- <u>again, we need gene sequence</u>

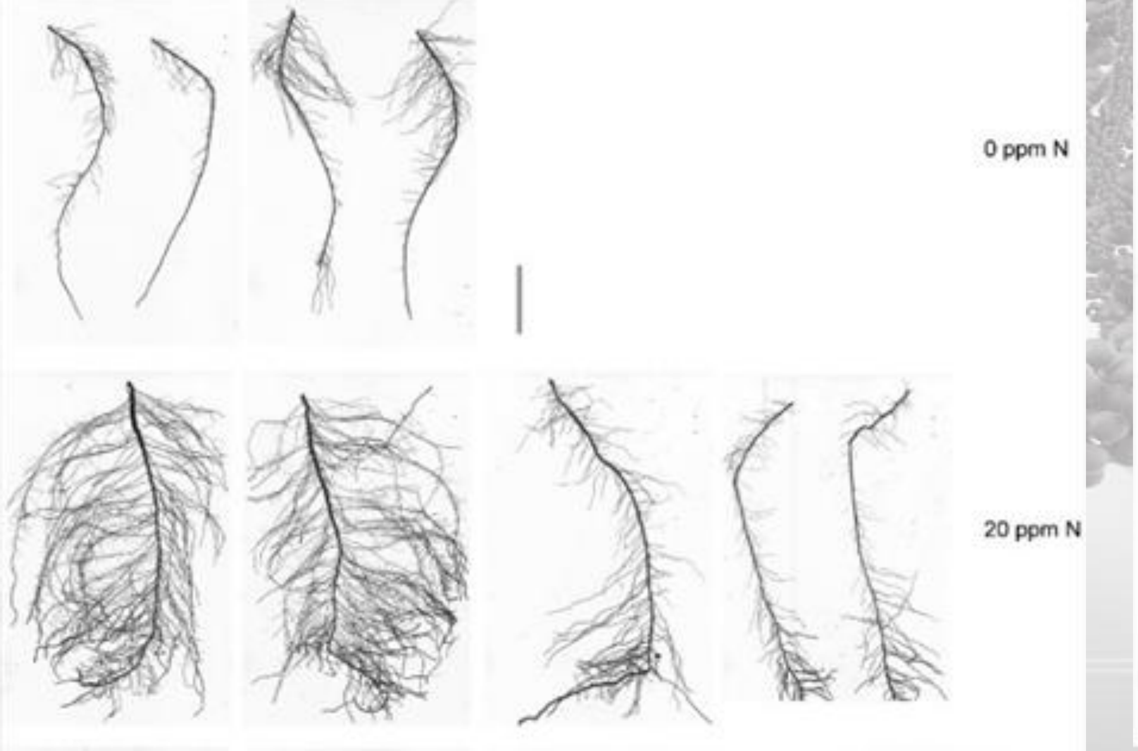


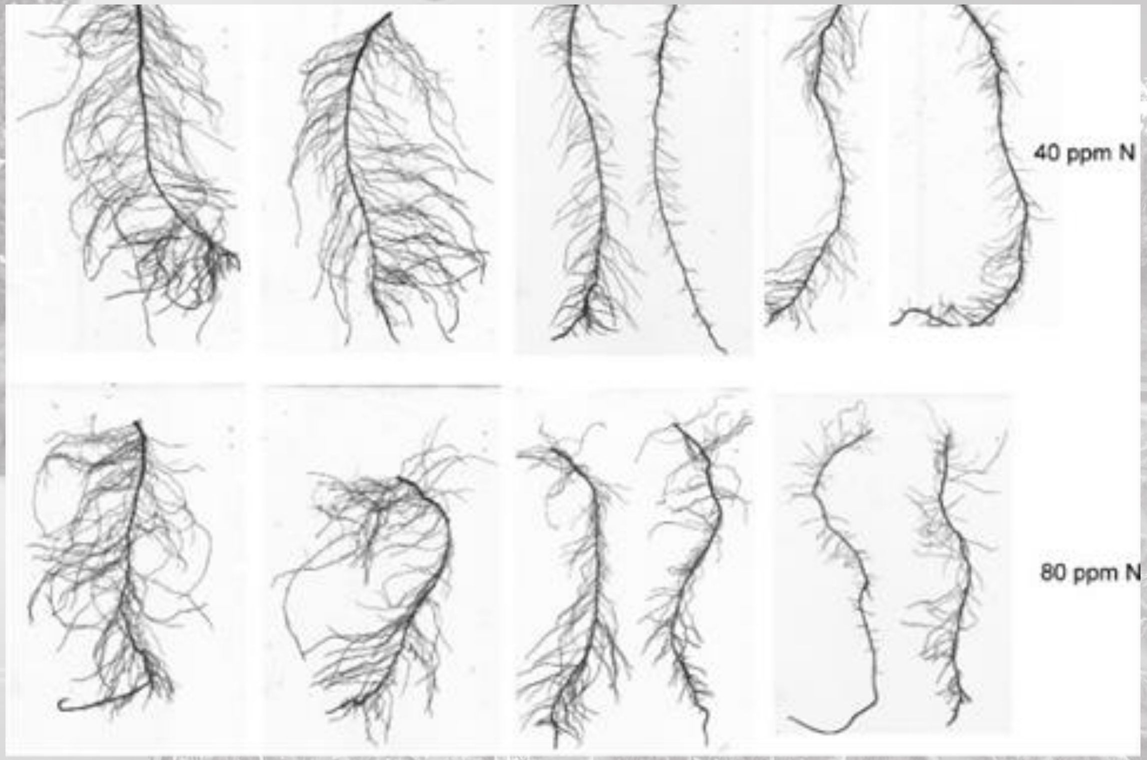
Optimization of chemical inputs and soil conditions



- focused on effects of inputs on yield and not on physiological properties of crops
- if crop physiology is optimum for nutrient absorption, minimum input would be needed
- knowledge of gene sequence for a specific physiological property would be extremely helpful

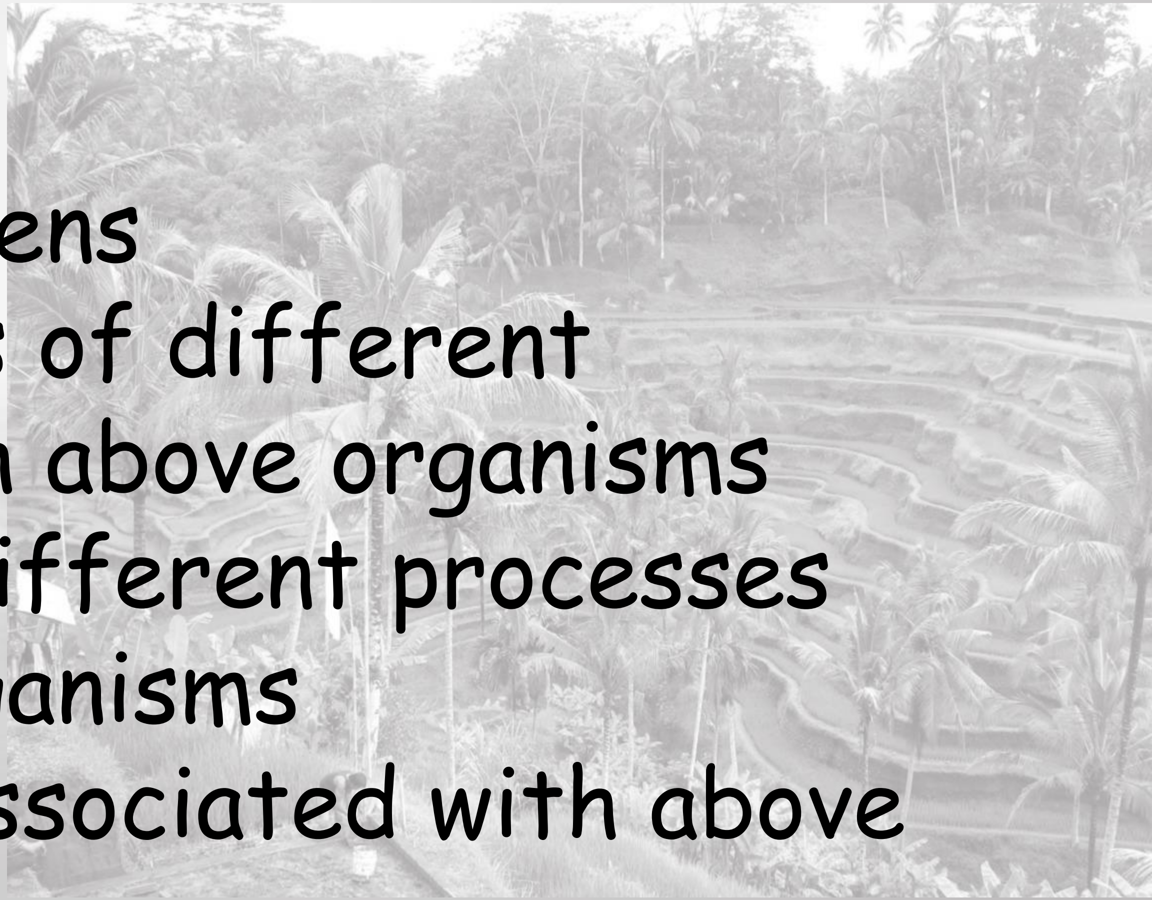
Optimization of chemical inputs and soil conditions



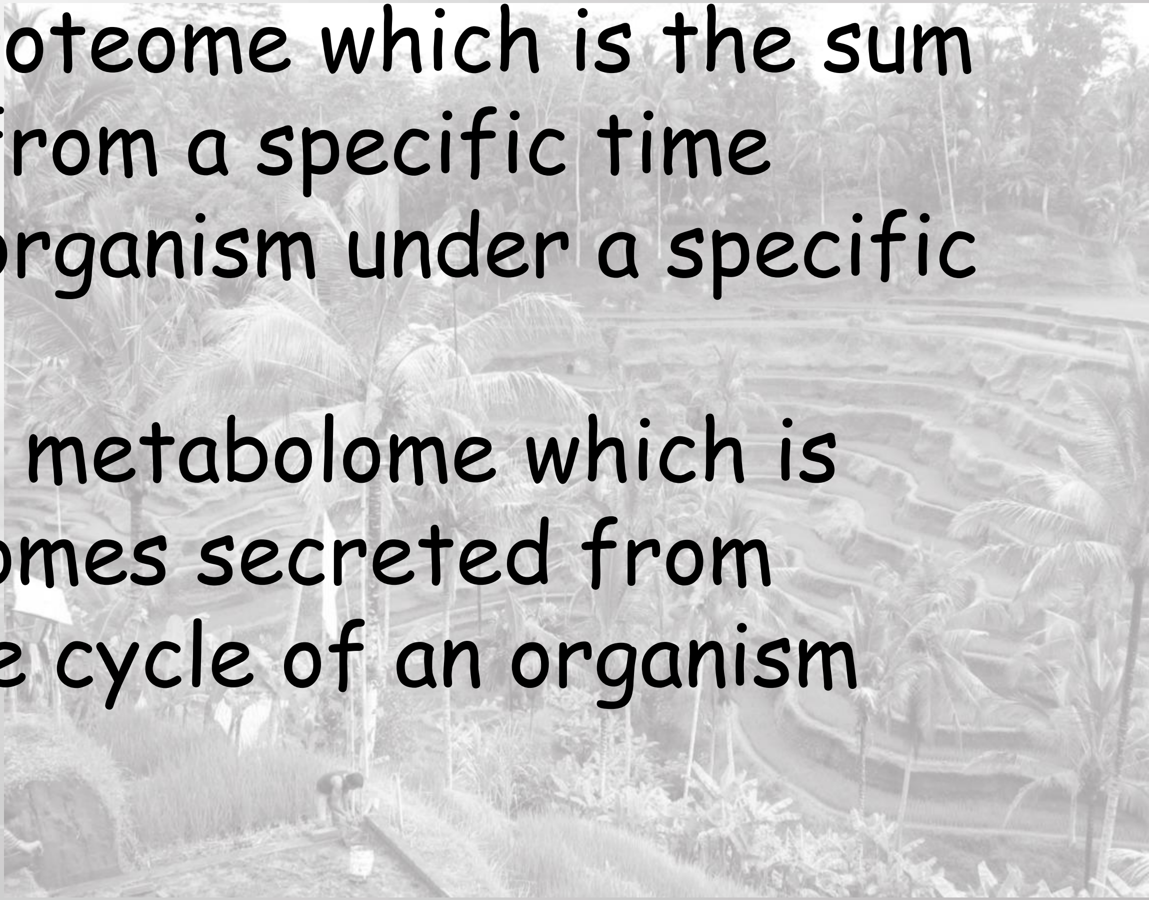


Current research in Genomics

- Genome sequences of different
  - animals including livestocks
  - plants including crops
  - plant and animal pests
  - plant and animal pathogens
- Transcriptomic sequences of different processes associated with above organisms
- Proteomic sequences of different processes associated with above organisms
- Metabolomic sequences associated with above

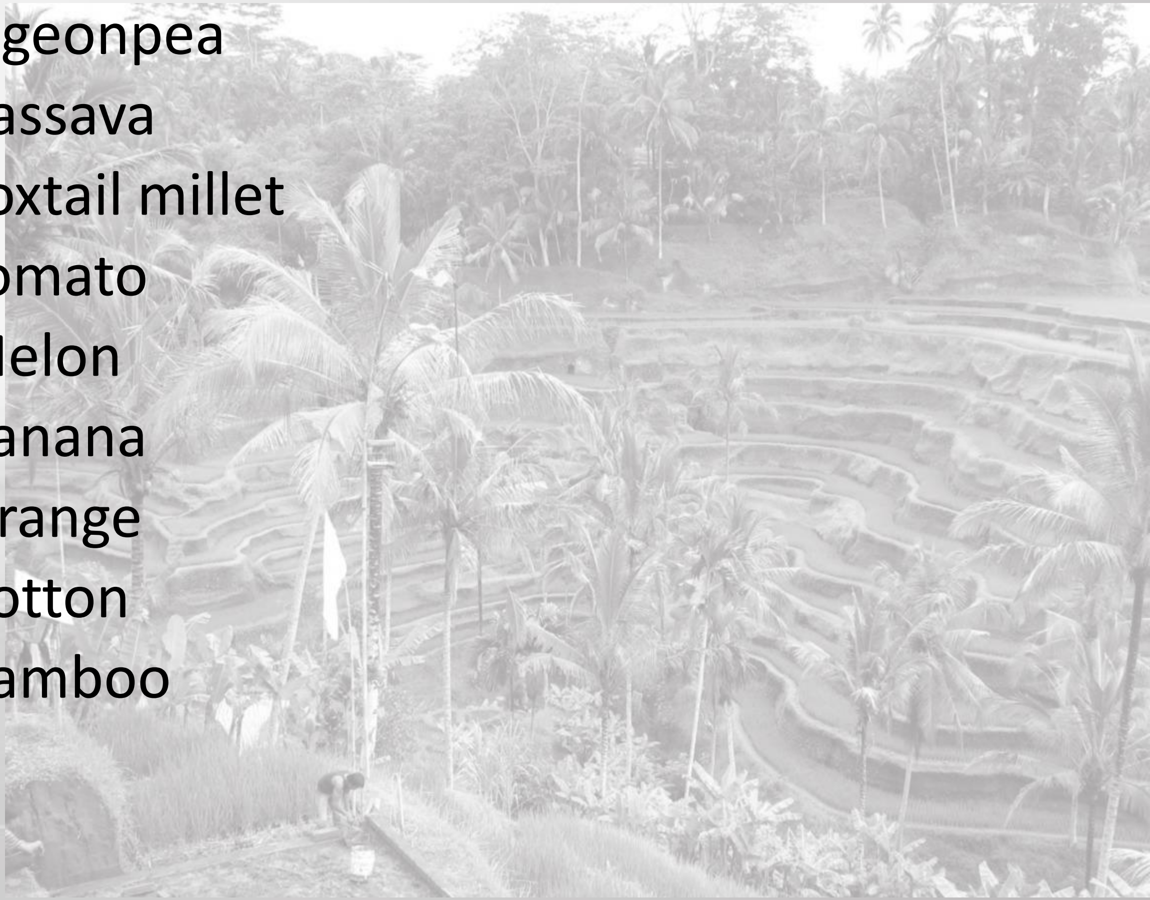


- Transcriptomics – study of the transcriptome which is the sum total of mRNA expression from a specific time point in the life cycle of an organism under a specific condition
- Proteomics – study of the proteome which is the sum total of protein expression from a specific time point in the life cycle of an organism under a specific condition
- Metabolomics – study of the metabolome which is the sum total of all metabolomes secreted from specific time point in the life cycle of an organism under a specific condition

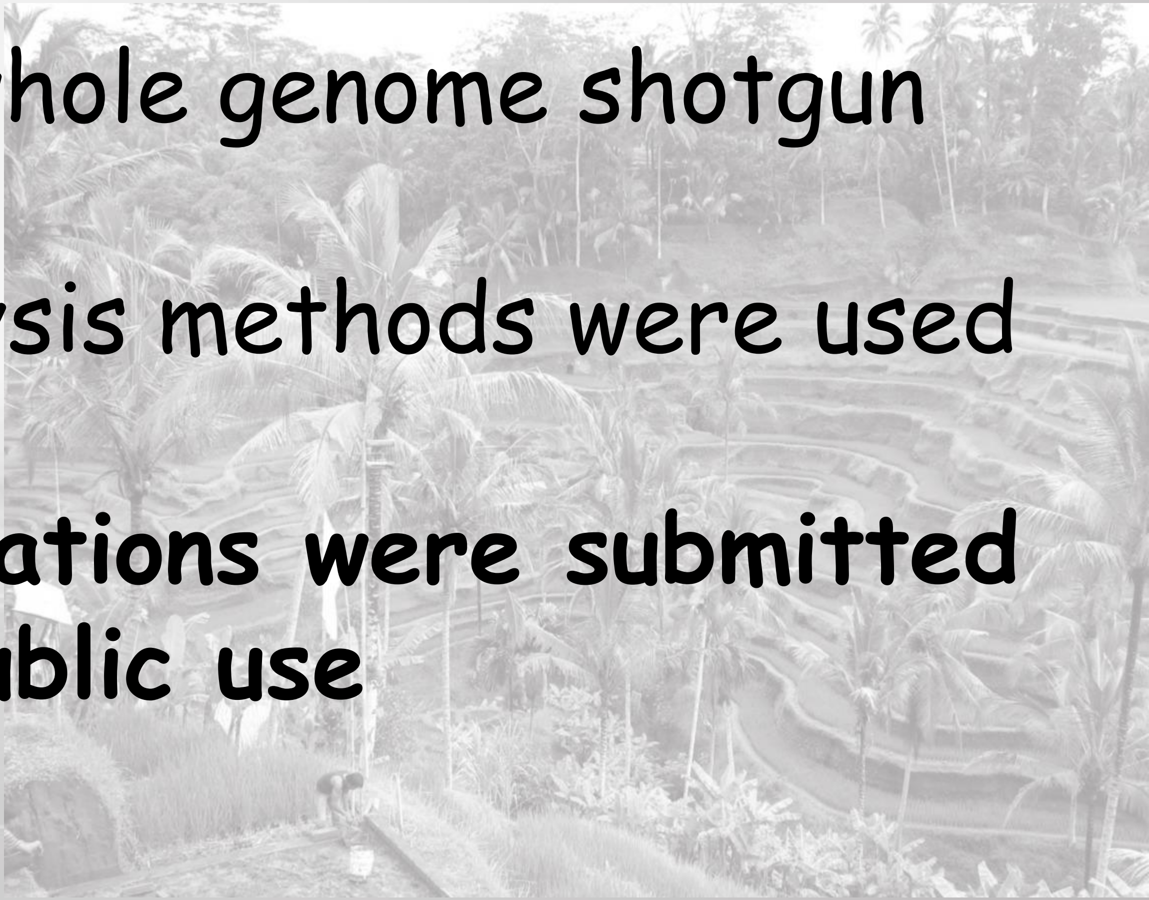


Crops with completed or partially completed genomes

- Rice
- Cottonwood
- Grapes
- Sorghum
- Maize
- Soybean
- Apple
- Strawberry
- Cacao
- Date palm
- Potato
- Chinese cabbage
- Alfalfa
- Pigeonpea
- Cassava
- Foxtail millet
- Tomato
- Melon
- Banana
- Orange
- Cotton
- Bamboo

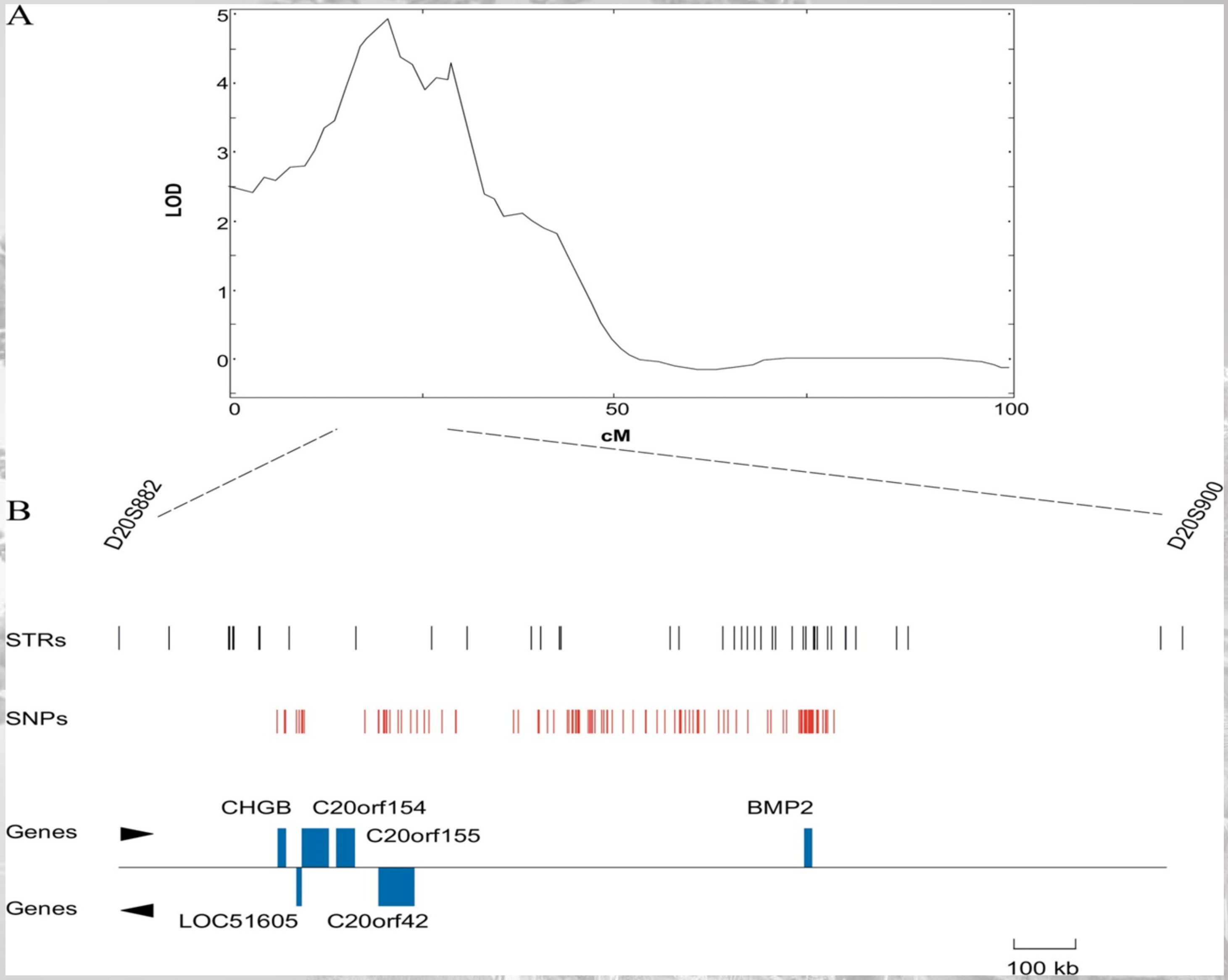


Crops with completed or partially completed genomes

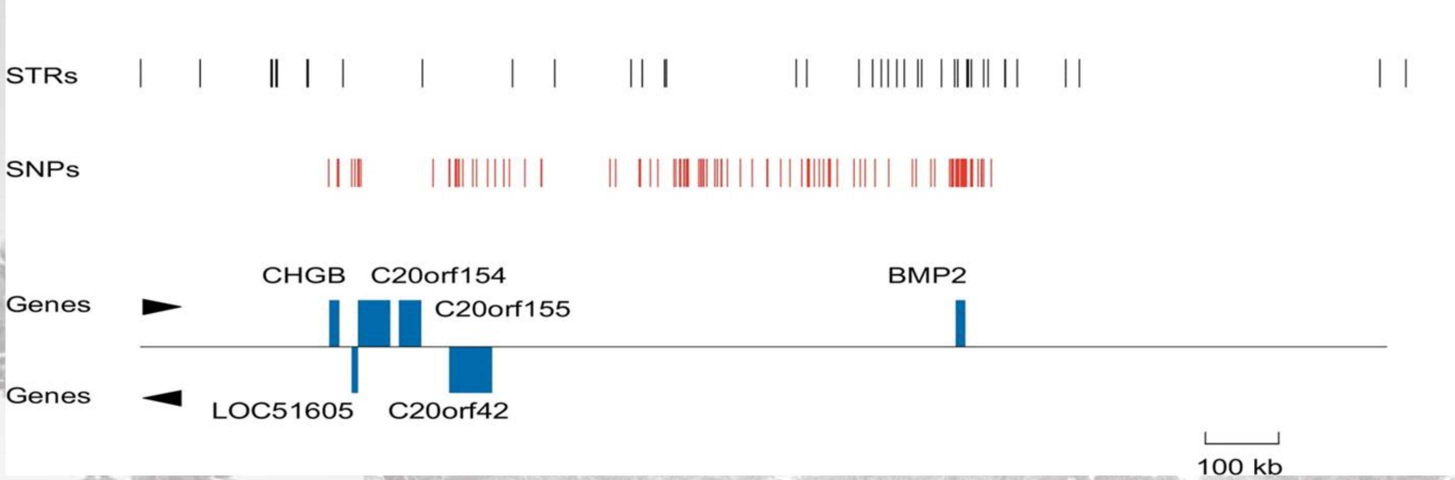


- Earlier projects used BAC libraries and dideoxy sequencing
- Later ones used NGS and whole genome shotgun techniques
- High-throughput data analysis methods were used for annotation
- **Data sequences and annotations were submitted to public databases for public use**

Data sequences and annotations may be used as precise molecular markers via PCR or qPCR



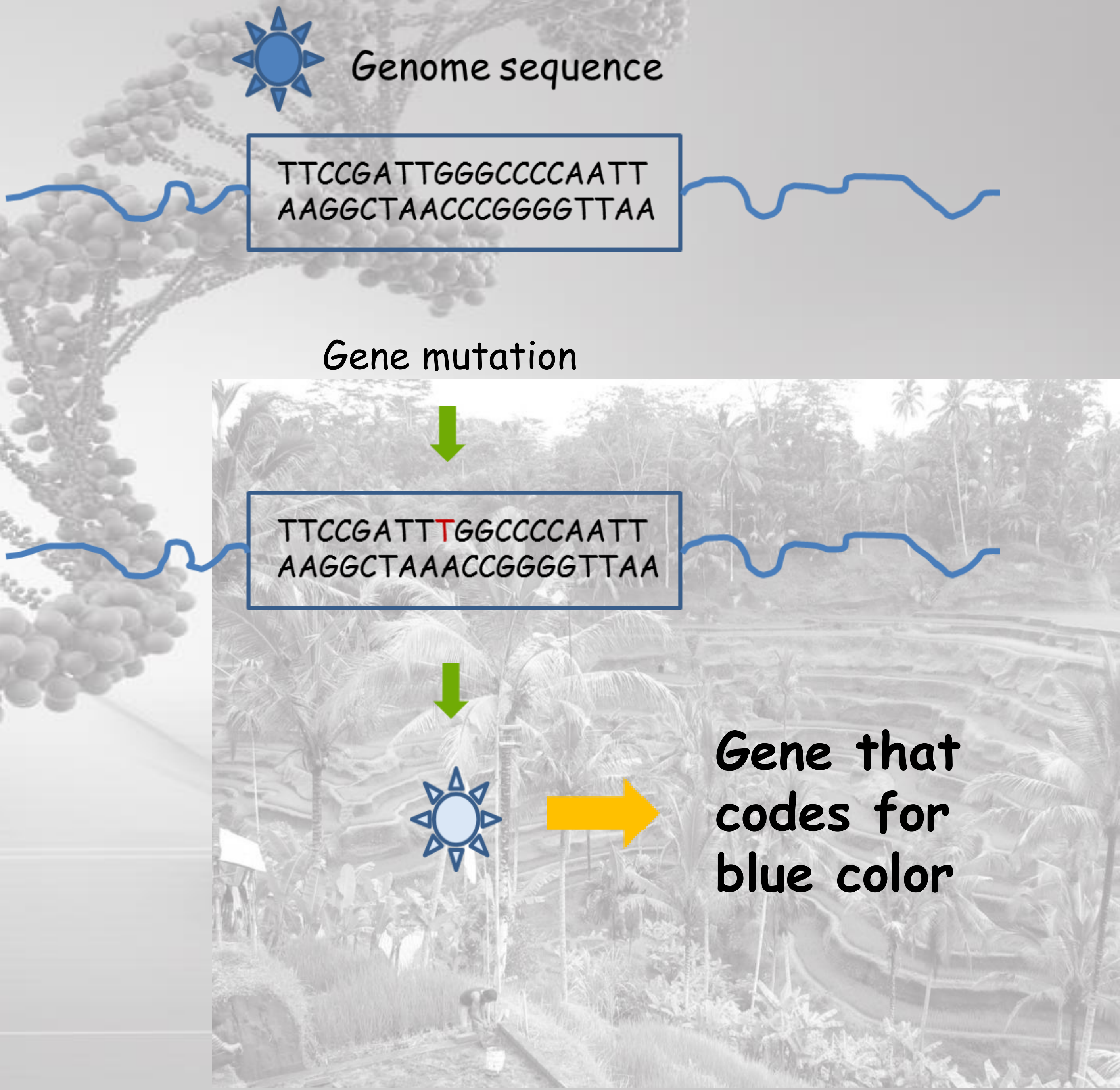
Data sequences and annotations may be used as precise molecular markers via PCR and sequencing



SNP1 (desirable trait)	CCCGTTAGGTAACCTTGGAAAGCGA
SNP2 (undesirable trait)	CCCGTTAGATAACCTTGGAAAGCGA

Data sequences and annotations may be used as target/s to focus on gene function studies (aka as reverse genetics)

- **Known gene sequences may be used as markers, transgenes or RNAi sequences**


Genome sequence

TTCCGATTGGGCCCCAATT
AAGGCTAACCCGGGGTTAA

Gene mutation

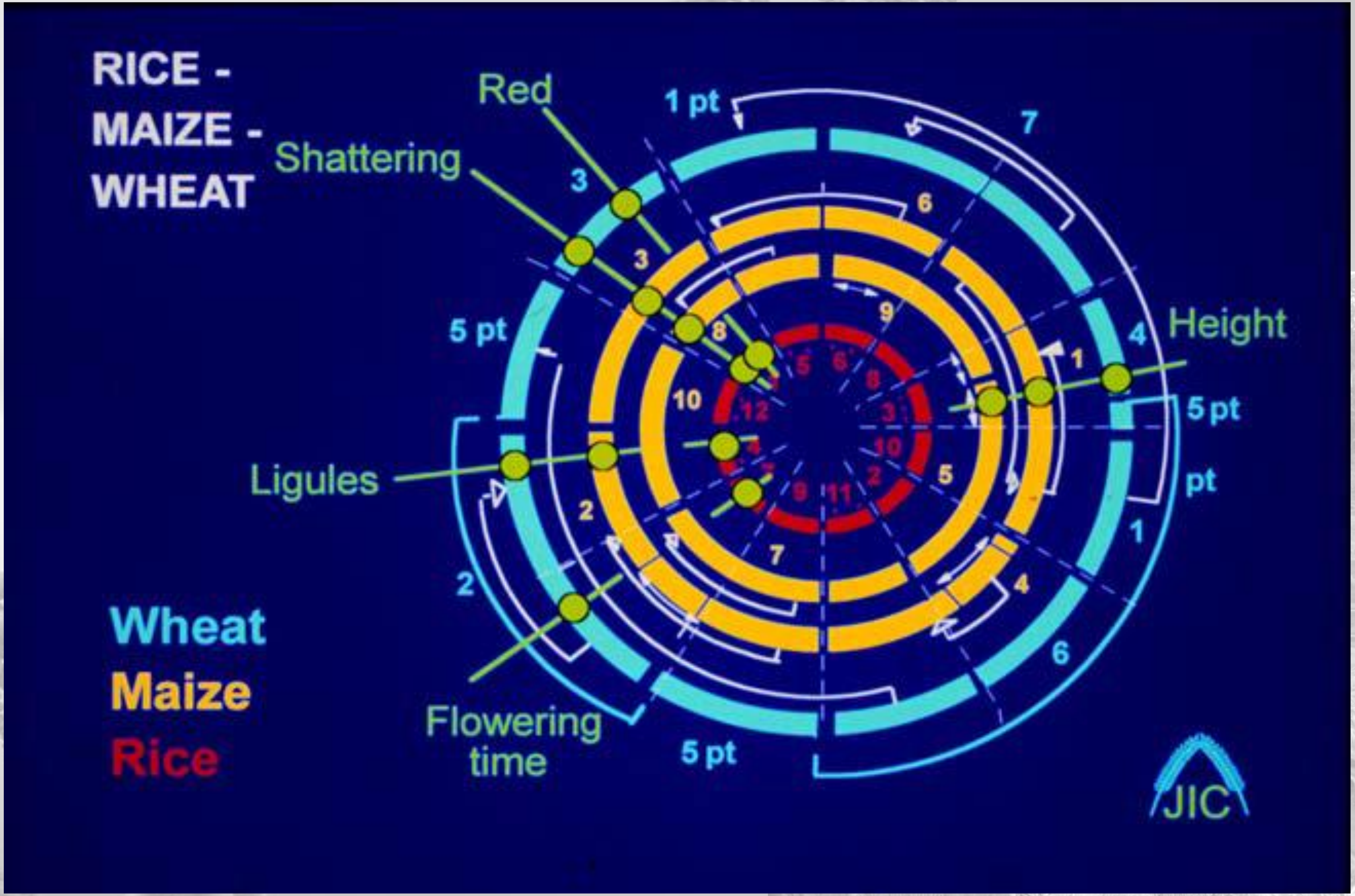
TTCCGATTTGGCCCCAATT
AAGGCTAAACCGGGGTTAA

Gene that codes for blue color

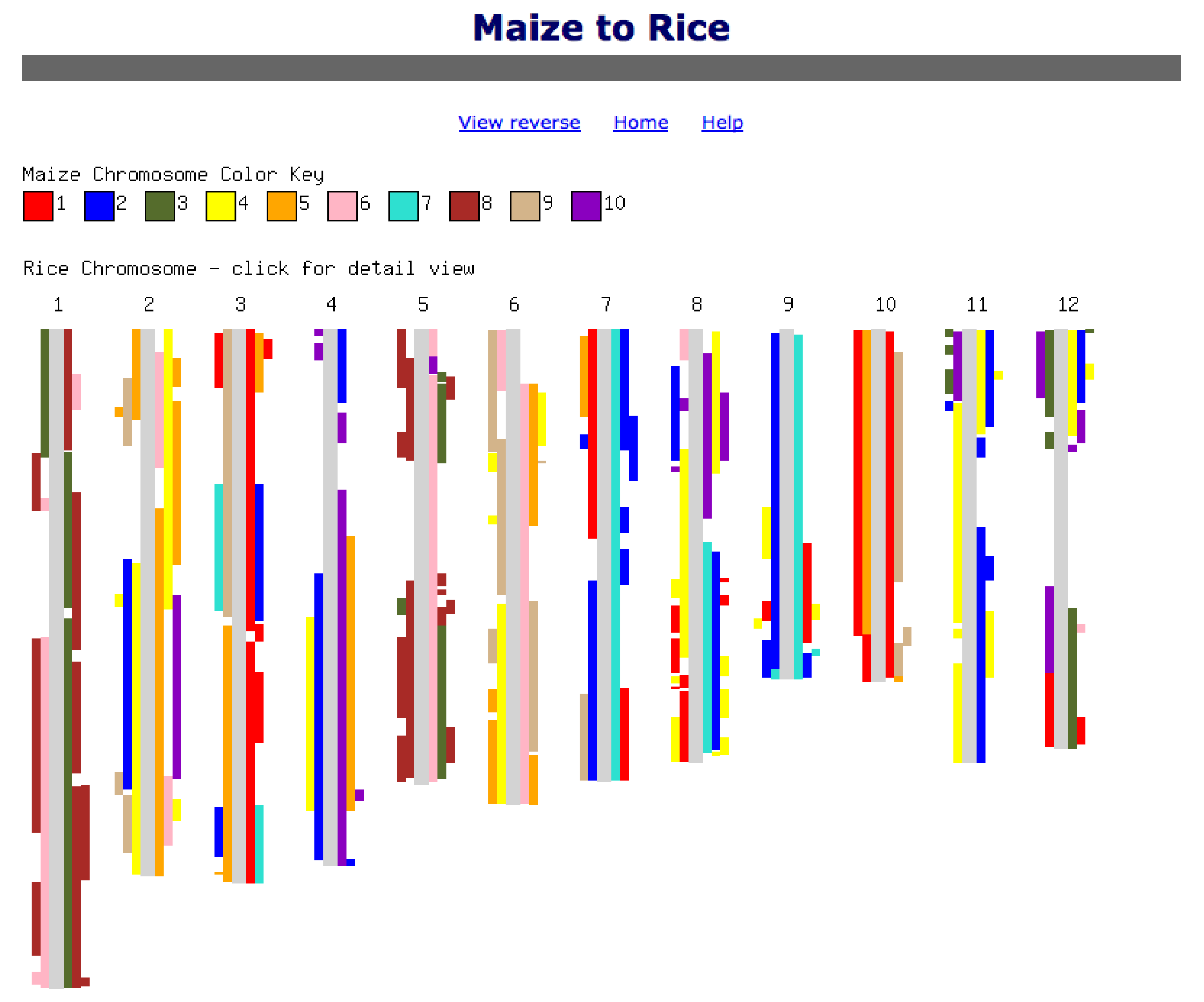
Characteristics of crop genomes

- Mostly huge
  - Number of genes almost the same
  - Repetitive elements content makes up the difference
  - Notorious for polyploidy
- However, gene order or synteny occurs

Synteny


RICE -
MAIZE -
WHEAT
Red
Shattering
Height
Ligules
Flowering time
1 pt
7
3
6
3
9
8
5 pt
4
1
5 pt
pt
10
5
1
2
7
4
2
6
5 pt
5
6
8
12
3
10
4
2
9
11
Wheat
Maize
Rice
JIC

Application of synteny



Maize to Rice

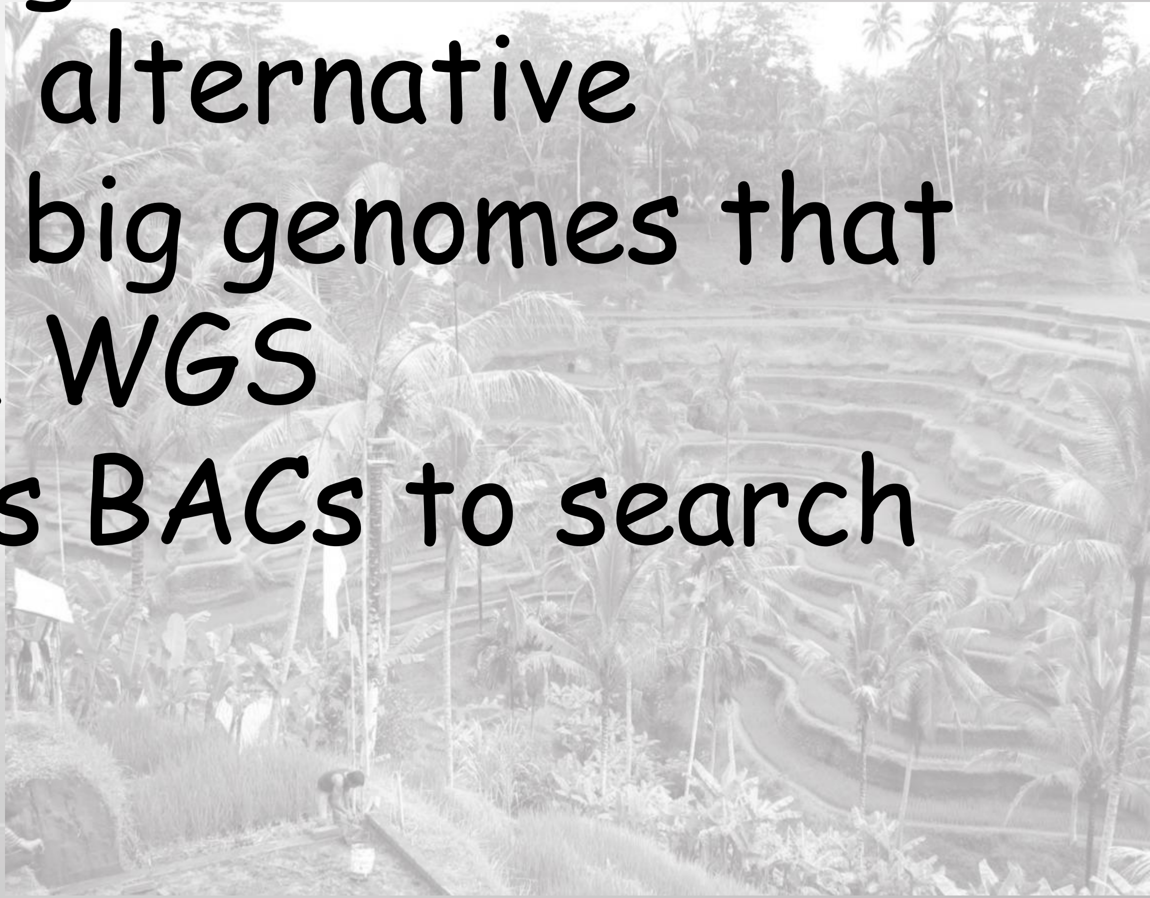
View reverse Home Help

Maize Chromosome Color Key

Rice Chromosome - click for detail view

Sequencing of crop genomes

- Greatest challenge is genome size
- Exome sequencing an alternative
- Use synteny to study big genomes that are not accessible via WGS
- Use resources such as BACs to search for genes of interest



Gene search in loblolly pine using BAC resources

Probes used

cell wall synthesis

- cellulose synthase
- cinnamyl alcohol dehyfrogenase
- expansin (alpha)
- fasciclin-like AGP
- glucomannan synthase
- glycoside hydrolase family 28
- Korrigan endoglucanase
- laccase
- lignin related
- phenylalanine ammonia lyase
- sucrose synthase
- UDP glucose pyrophosphorylase
- UGP mannose pyrophosphorylase
- vein patterning

transcription factors

- assymetric leaves-like; LOB domain transcription factor
- brassinosteroid responsive transcription factor
- brassinosteroid-insensitive; shaggy-like protein kinase
- bZIP transcription factor
- HD Zip III
- Kanadi, Myb-related protein
- KNOX homeobox transcription factor
- LIM domain protein
- Myb family transcription factor
- NAC domain
- R2R3-MYB transcription factor
- transcription factor WRKY1

disease resistance

Gene search in loblolly pine using BAC resources

Masked repeats using Repeatmasker

→

Retrotransposons the most abundant repeats

↓

BLAST, FGenesh, Augustus and GeneMark to predict genes and Artemis to view predictions

Gene search in loblolly pine using BAC resources

Masked repeats using Repeatmasker

Retrotransposons the most abundant repeats

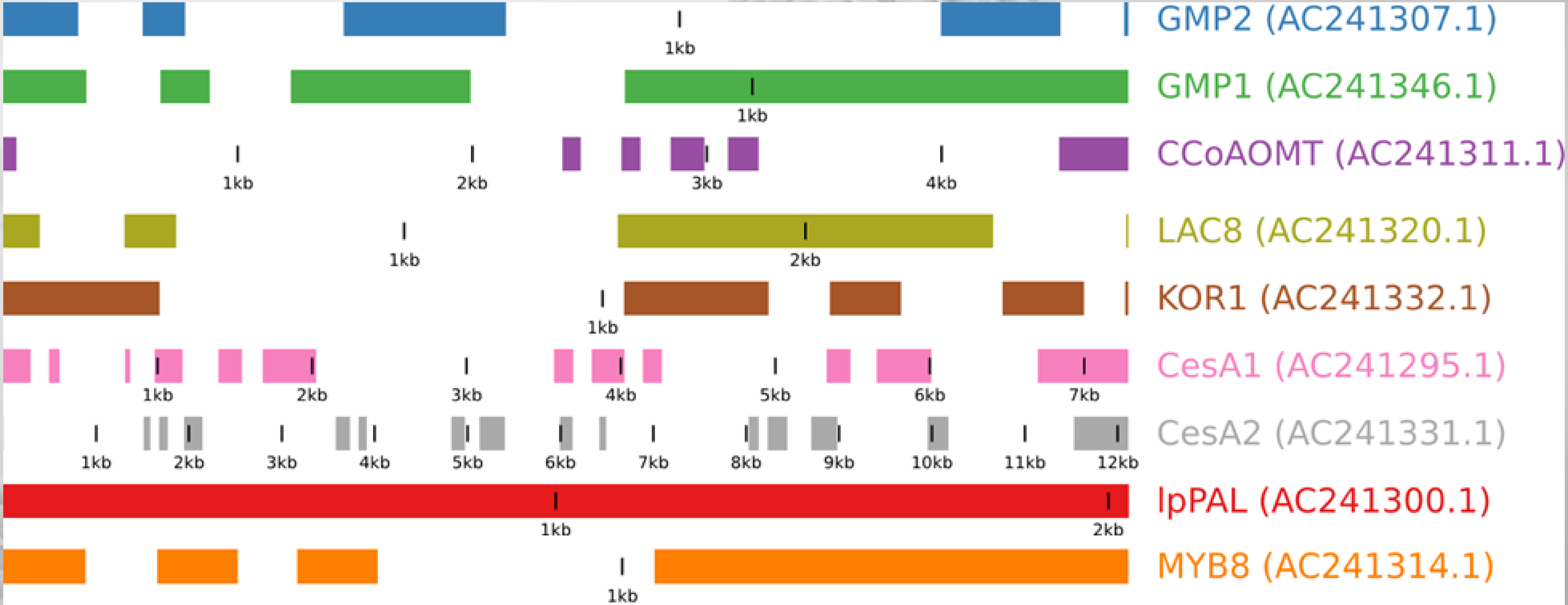
BLAST, FGenesh, Augustus and GeneMark to predict genes and Artemis to view predictions

Annotation of 114 gene models

9 predicted genes

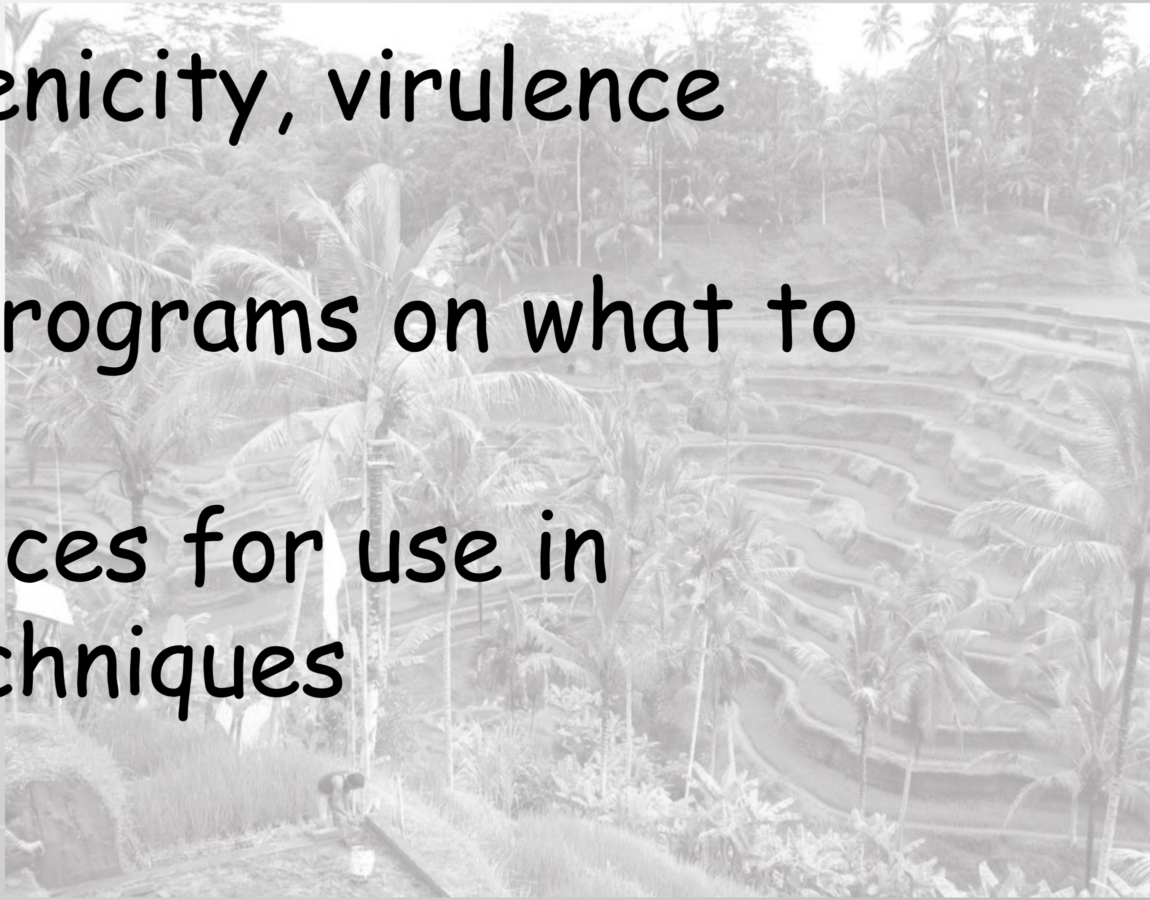
105 pseudogenes

Predicted Gene Models



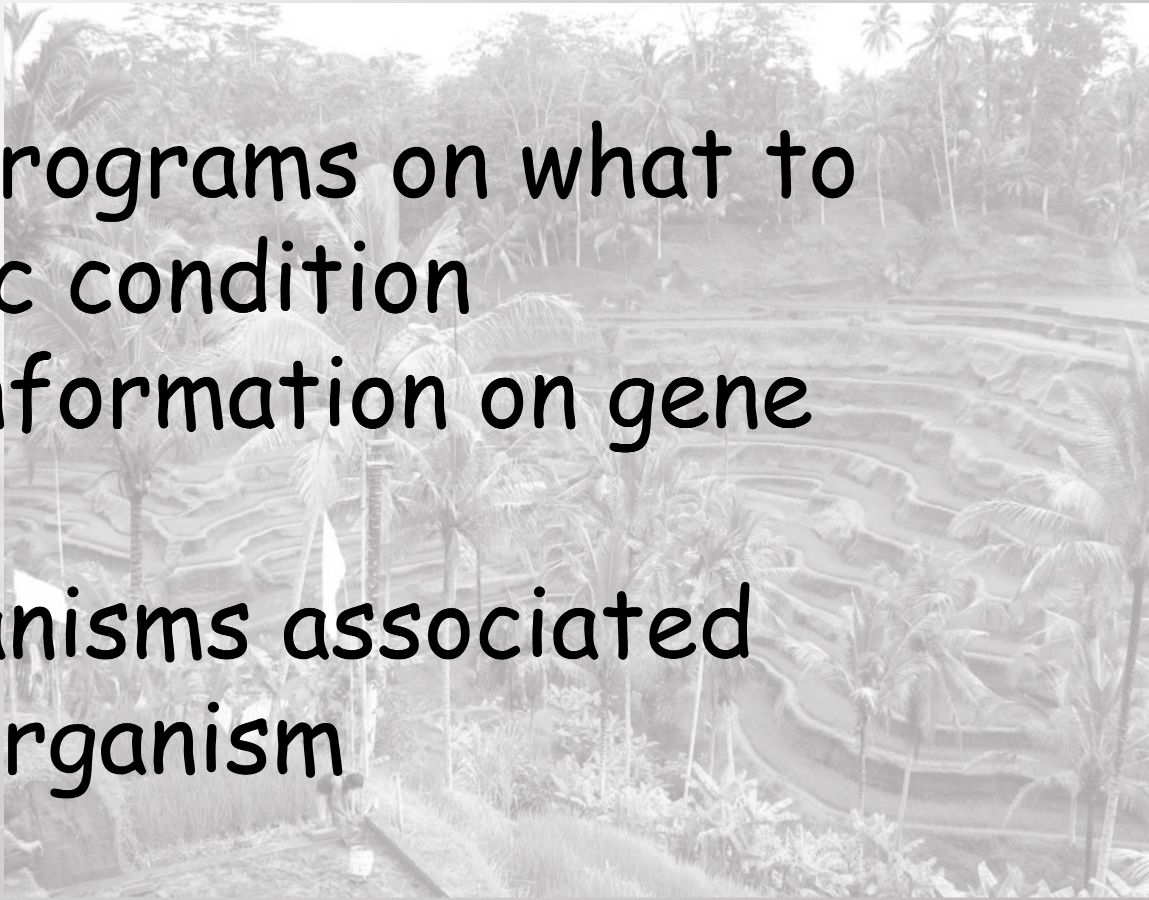
Benefit from pathogen and pest genome sequences

- Provide clues on pathogenicity, virulence and/or infectivity
  - May assist breeding programs on what to select
- May afford gene sequences for use in transgenic and RNAi techniques



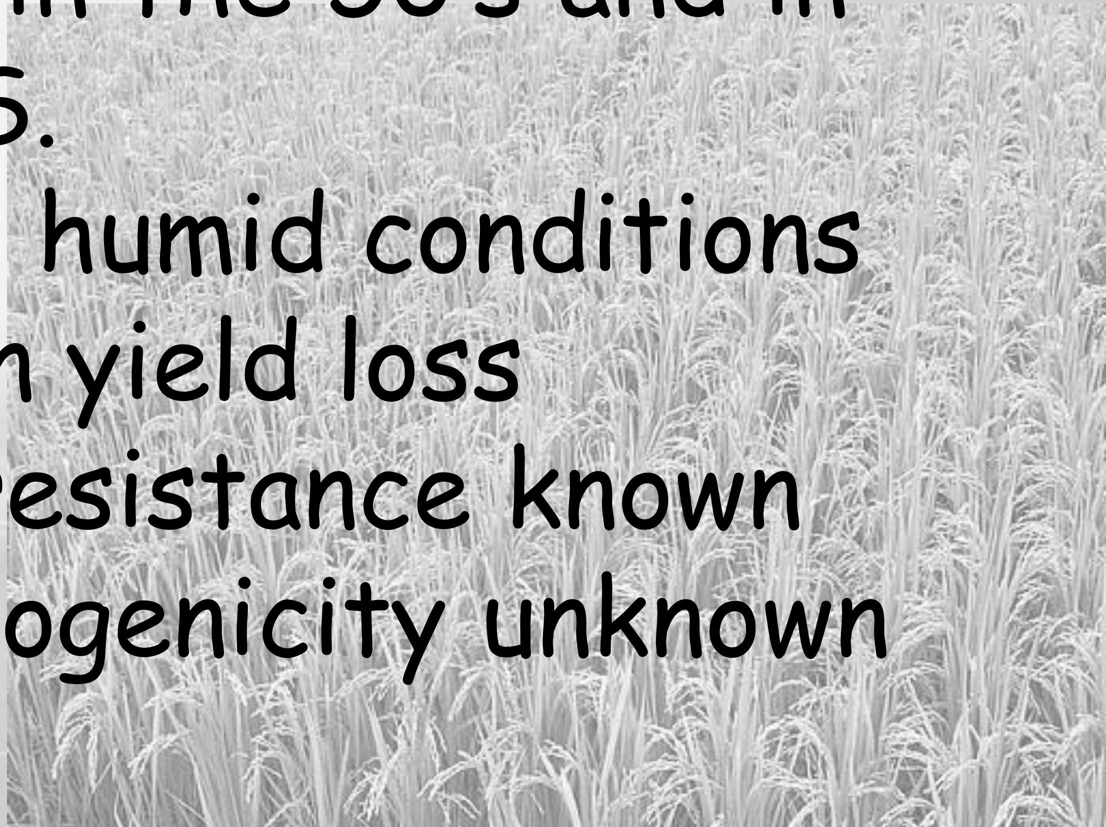
Benefit from transcriptomic/ proteomic/metabolomic research

- Allow for a better understanding of the biology of an organism
    - May assist breeding programs on what to select under a specific condition
- May afford additional information on gene and protein expression
- Aid in identifying mechanisms associated with a process for the organism



Bacterial panicle blight

Bacterial panicle blight (BPB)



- Caused by *Burkholderia glumae*
- First observed in Asia in the 50's and in the late 90's in the U.S.
- Aggravated by hot and humid conditions
- May cause up to 70% in yield loss
- No reliable source of resistance known
- Disease cycle and pathogenicity unknown

Searching for mechanism of resistance

CL161
Resistant rice seedlings

and

CL151
Susceptible rice seedlings

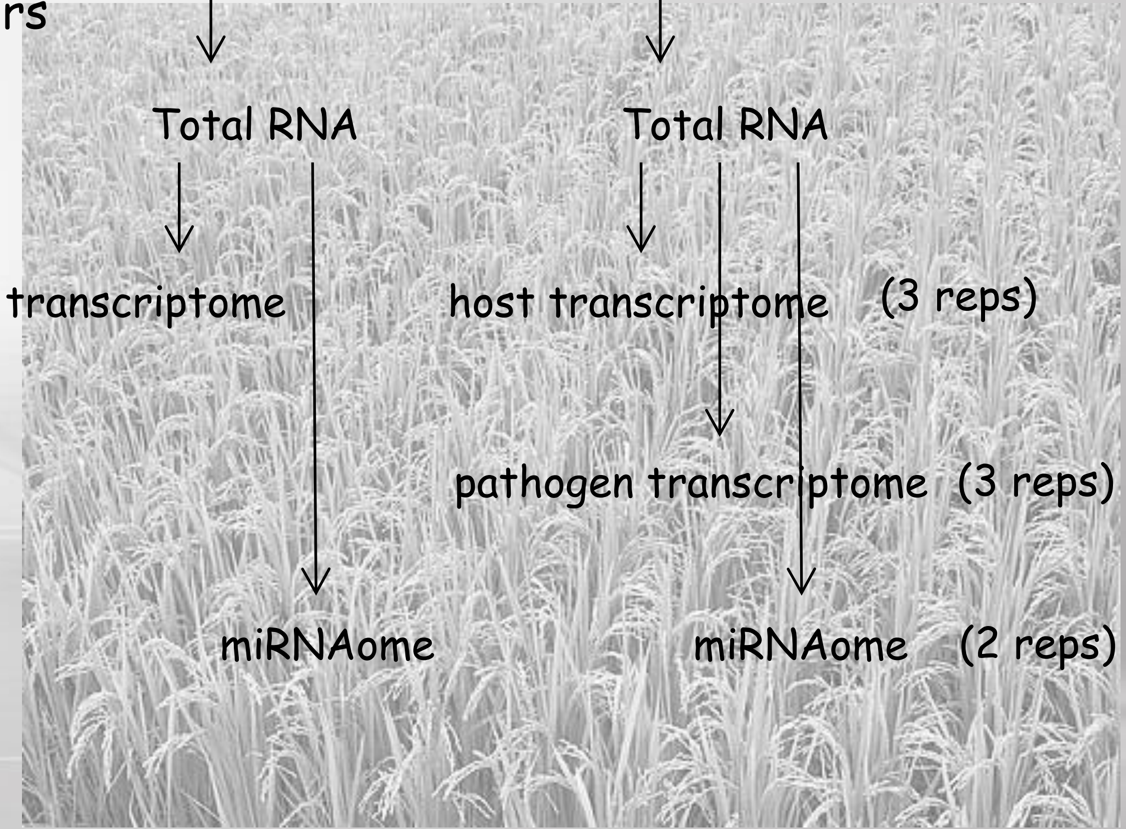
sample after 48 hours

Water inoculated → Total RNA → host transcriptome; miRNAome

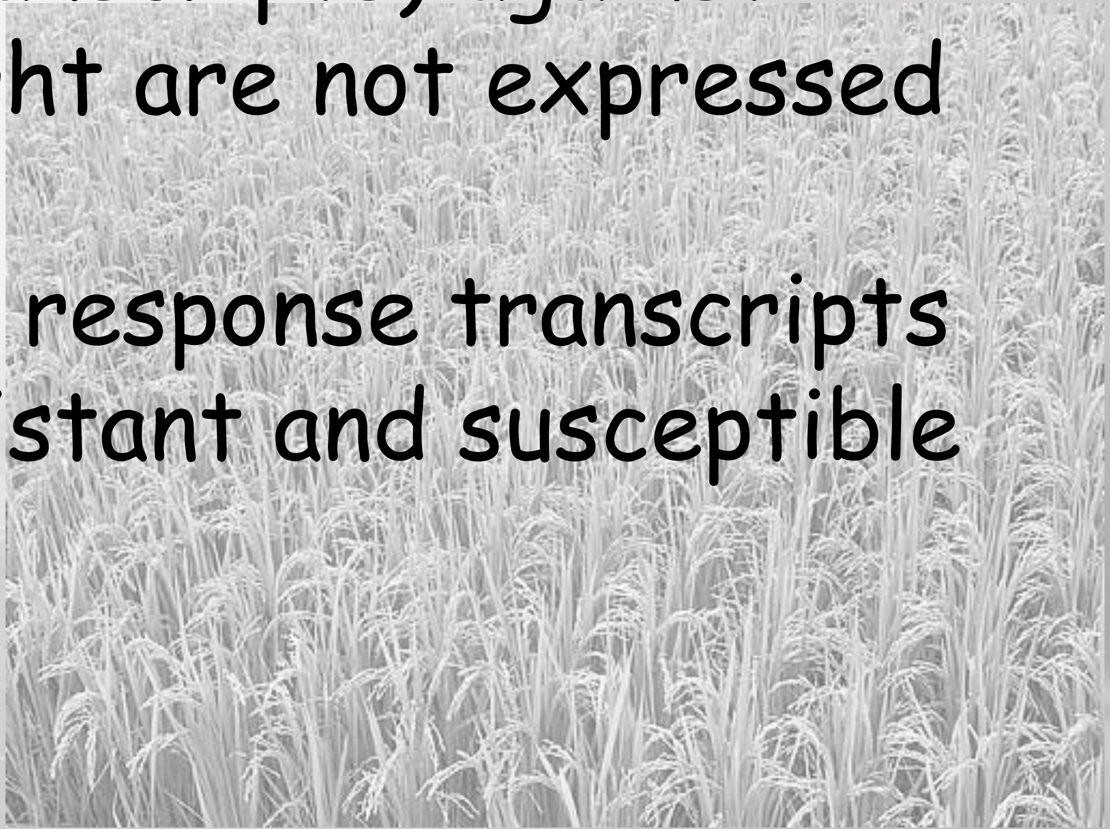
B. glumae inoculated → Total RNA → host transcriptome; pathogen transcriptome; miRNAome

Water inoculated → Total RNA → host transcriptome; miRNAome

B. glumae inoculated → Total RNA → host transcriptome (3 reps); pathogen transcriptome (3 reps); miRNAome (2 reps)



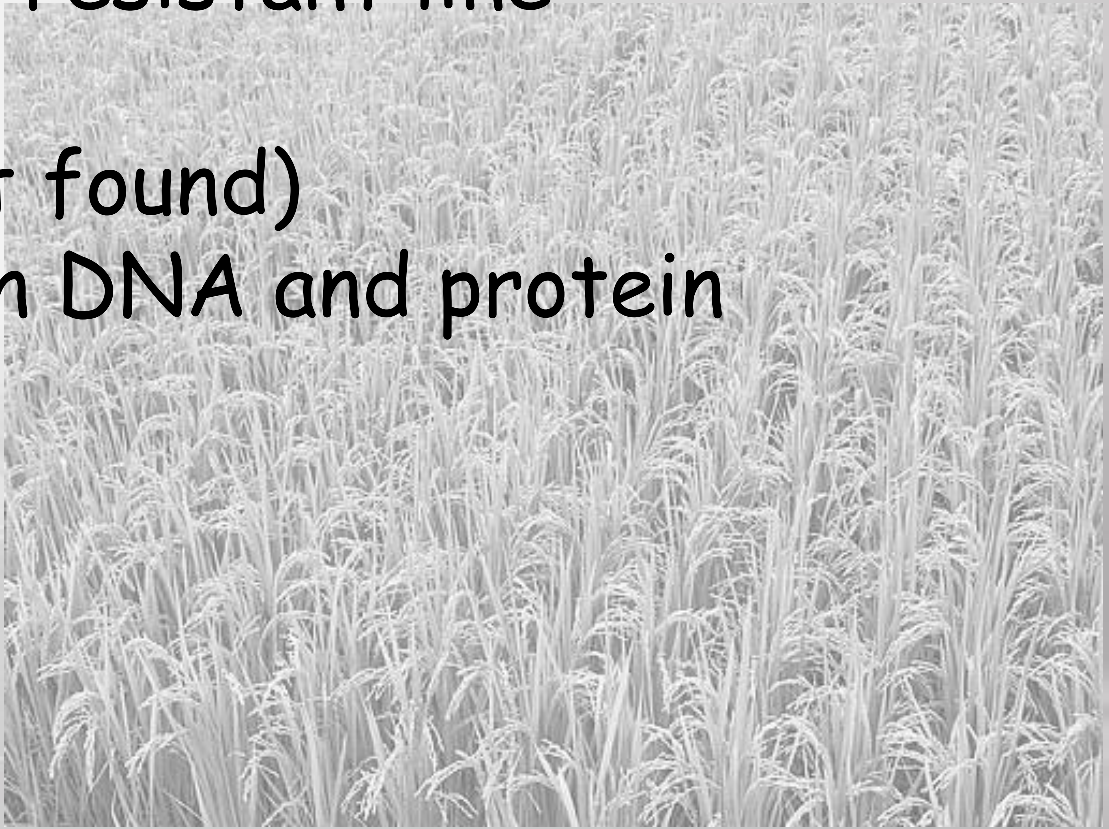
A cursory look at the DE transcripts



- Known resistance genes (transcripts) against rice blast and bacterial blight are not expressed at the inoculated condition
- Different kinds of defense response transcripts are up-regulated in the resistant and susceptible genotypes

PIF-like ORF1

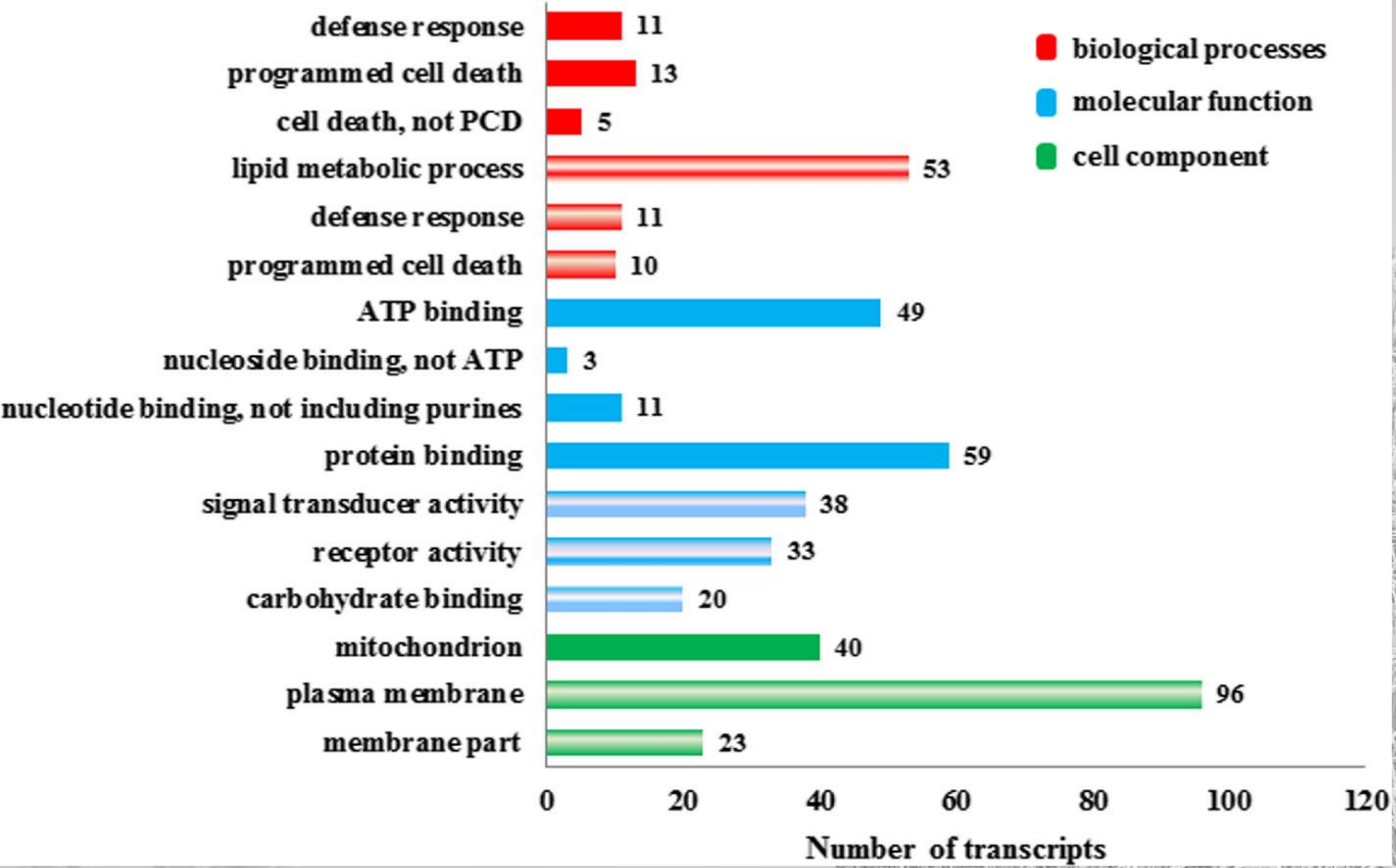
- P *instability factor*
- Class 2 transposable element
- Constitutively expressed in resistant line
- PIF has 2 ORFs
  - ORF2 is transposase (not found)
  - ORF1 probably involved in DNA and protein binding



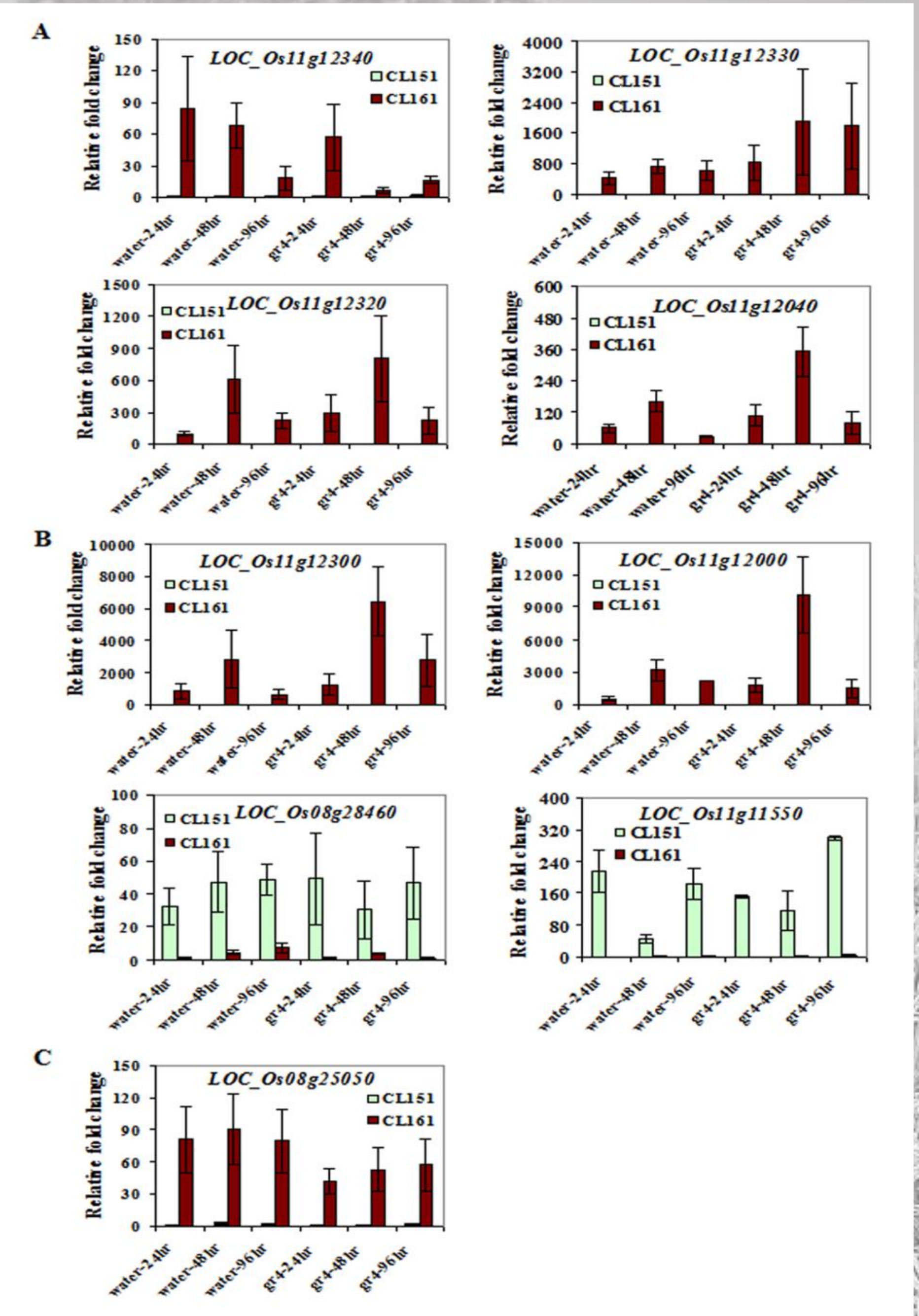
NBS-LRR resistance genes

- Ancient
- Involved in pathogen recognition
- Present in large numbers
  - Occurs in clusters (most cases)
  - A lot are pseudogenes (illegitimate recombination)
- Usually negatively regulated
  - Activated by ATP

Gene Ontology annotation of differentially expressed transcripts

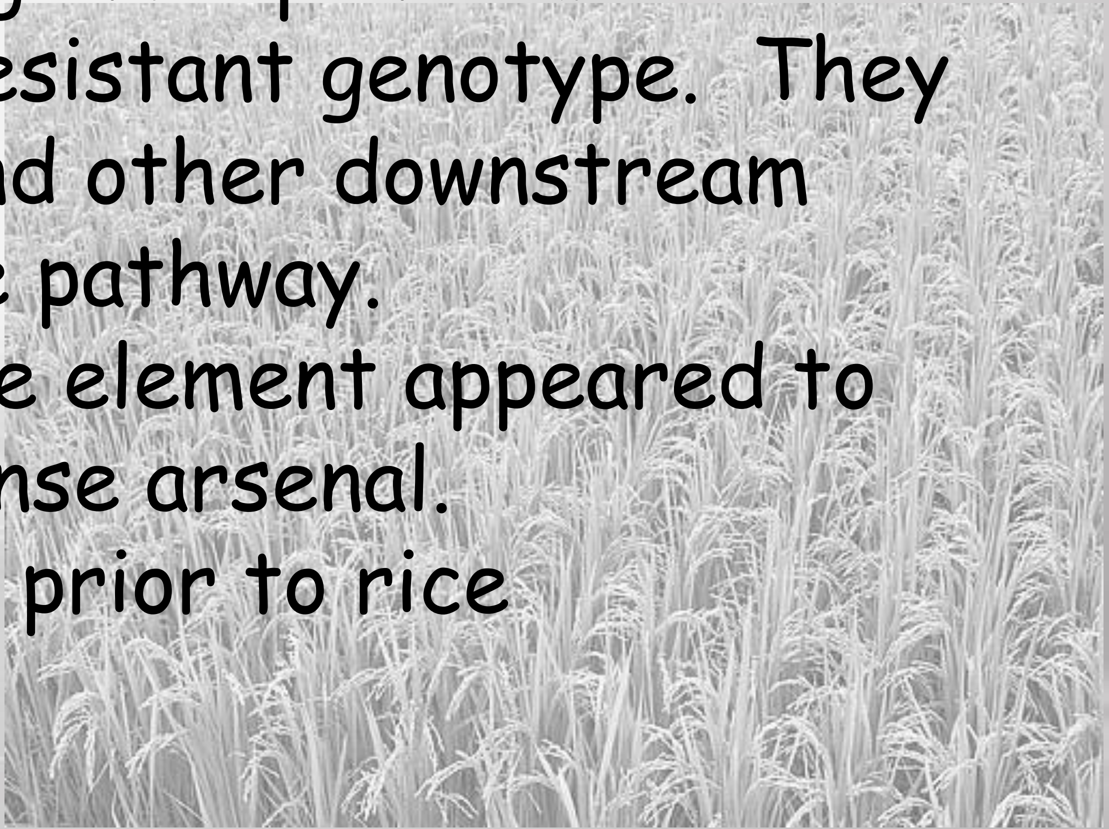


Quantitative RT-PCR Validation



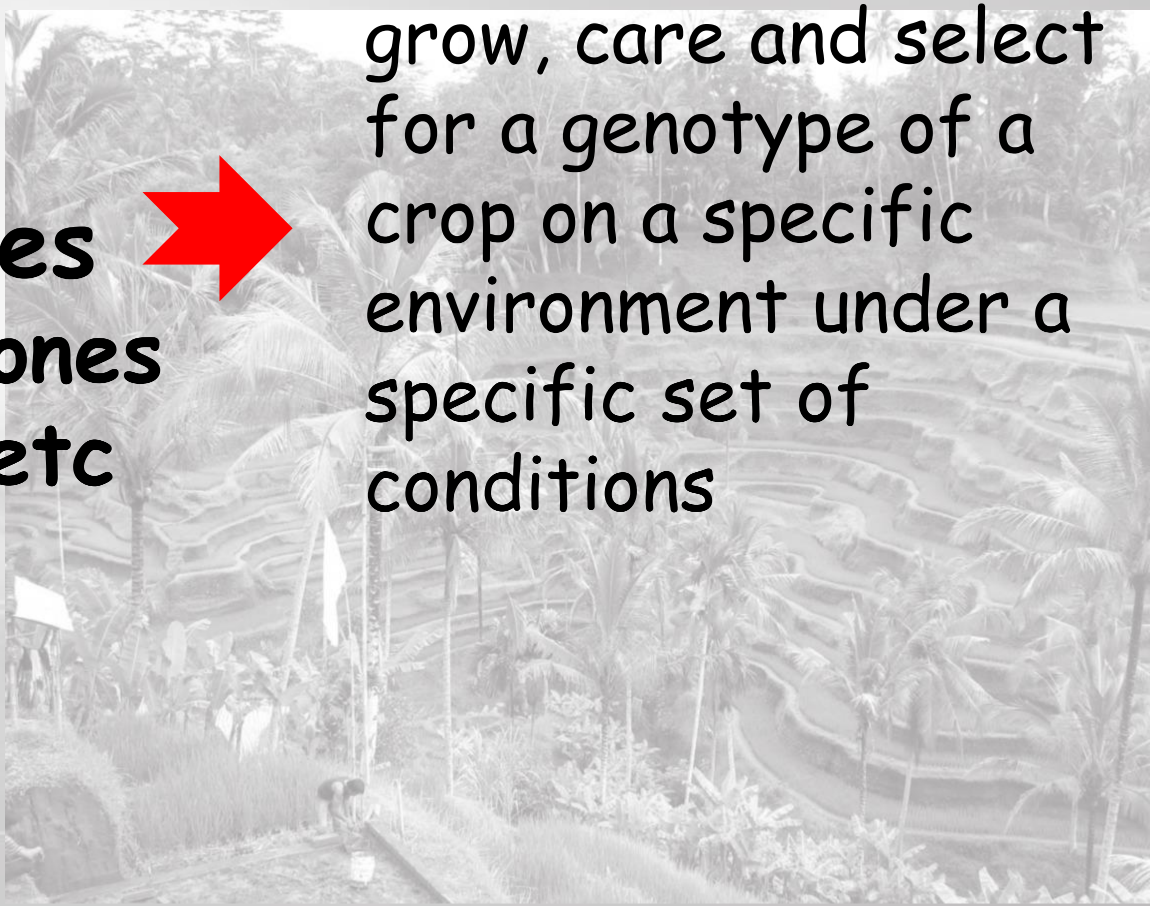
Conclusion (transcriptomic section):

1. Known resistance genes against rice blast and bacterial blight are not expressed in the resistant genotype; some not even in the water control.
2. Transcripts involved in pathogen response are constitutively expressed in resistant genotype. They include activation partners and other downstream components of the resistance pathway.
3. A once part of a transposable element appeared to have been co opted as a defense arsenal.
4. Resistance may have existed prior to rice domestication



Public Databases

- ➢ cell phones tablets, etc



Details on how to grow, care and select for a genotype of a crop on a specific environment under a specific set of conditions

ACKNOWLEDMENT

University of the Philippines
National Institute of Molecular Biology and Biochemistry
Department of Science and Technology
Philippine Council for Industry, Energy and Emerging Technology Research and Development

Dr. Dan Peterson
Dr. Dana Nelson
Dr. Jeff Dean
Dr. Joe Nairn
Dr. Ron Sederof
Dr. Dinum Perera
Ms. Seval Oskan
Mr. Ben Bartlett

Dr. Shien Lu
Dr. Teresia Buza
Dr. Chuan-Yu Hsu
Mr. Kurt Showmaker
Mr. Peng Deng
Dr. Philippe Chouvarine
Mr. Tony Arick