

# Pan-Millet Genomics and Genetics

**Jeff Bennetzen**  
**University of Georgia**



# Comparative Genomics and Genetics



# Comparative Genomics and Genetics

- ★ **What is it?**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**

- **Synergy from the multi-species approach (models!)**



# Comparative Genomics and Genetics

## ★ **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## ★ **Why do it?**

- **Synergy from the multi-species approach (models!)**
- **Empowers minor (e.g., orphan) species**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**

- **Synergy from the multi-species approach (models!)**
- **Empowers minor (e.g., orphan) species**
- **Illuminates significance and novelty by the degree of gene conservation**



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**

- **Synergy from the multi-species approach (models!)**
- **Empowers minor (e.g., orphan) species**
- **Illuminates significance and novelty by the degree of gene conservation**

## \* **How is it done?**



# Comparative Genomics and Genetics

## ★ What is it?

- Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function
- Lots of DNA sequence analysis

## ★ Why do it?

- Synergy from the multi-species approach (models!)
- Empowers minor (e.g., orphan) species
- Illuminates significance and novelty by the degree of gene conservation

## ★ How is it done?

- Comparisons of sequences (genomic, expressed)



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**

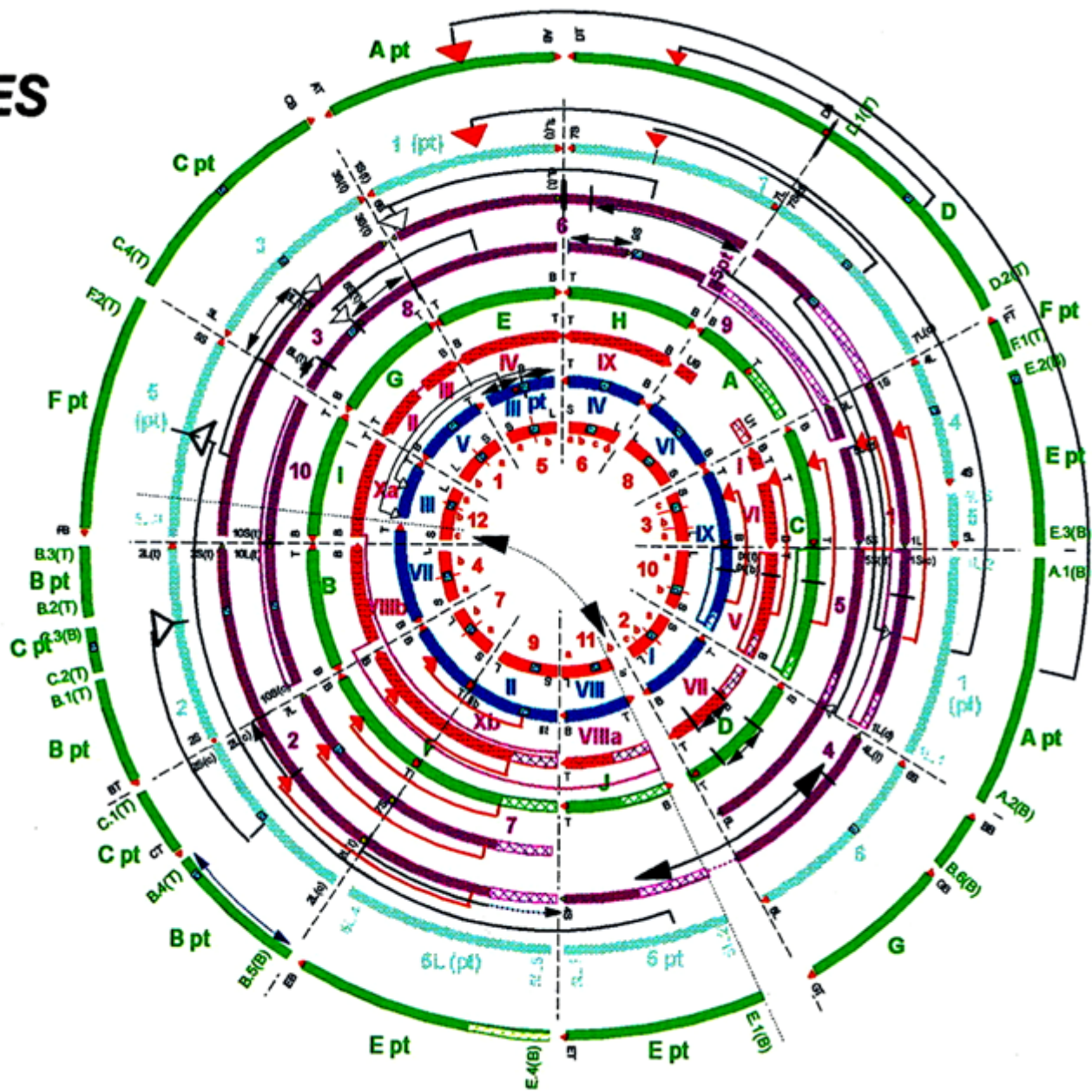
- **Synergy from the multi-species approach (models!)**
- **Empowers minor (e.g., orphan) species**
- **Illuminates significance and novelty by the degree of gene conservation**

## \* **How is it done?**

- **Comparisons of sequences (genomic, expressed)**
- **Comparisons of genetic maps**



# GRASS GENOMES



- Oats
- Triticeae
- Maize
- Sorghum
- Sugar cane
- Foxtail millet
- Rice



# Comparative Genomics and Genetics

## \* **What is it?**

- **Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function**
- **Lots of DNA sequence analysis**

## \* **Why do it?**

- **Synergy from the multi-species approach (models!)**
- **Empowers minor (e.g., orphan) species**
- **Illuminates significance and novelty by the degree of gene conservation**

## \* **How is it done?**

- **Comparisons of sequences (genomic, expressed)**
- **Comparisons of genetic maps**



# Comparative Genomics and Genetics

## ★ What is it?

- Cross-haplotype (e.g., cross-species) comparisons of gene content, gene structure and gene function
- Lots of DNA sequence analysis

## ★ Why do it?

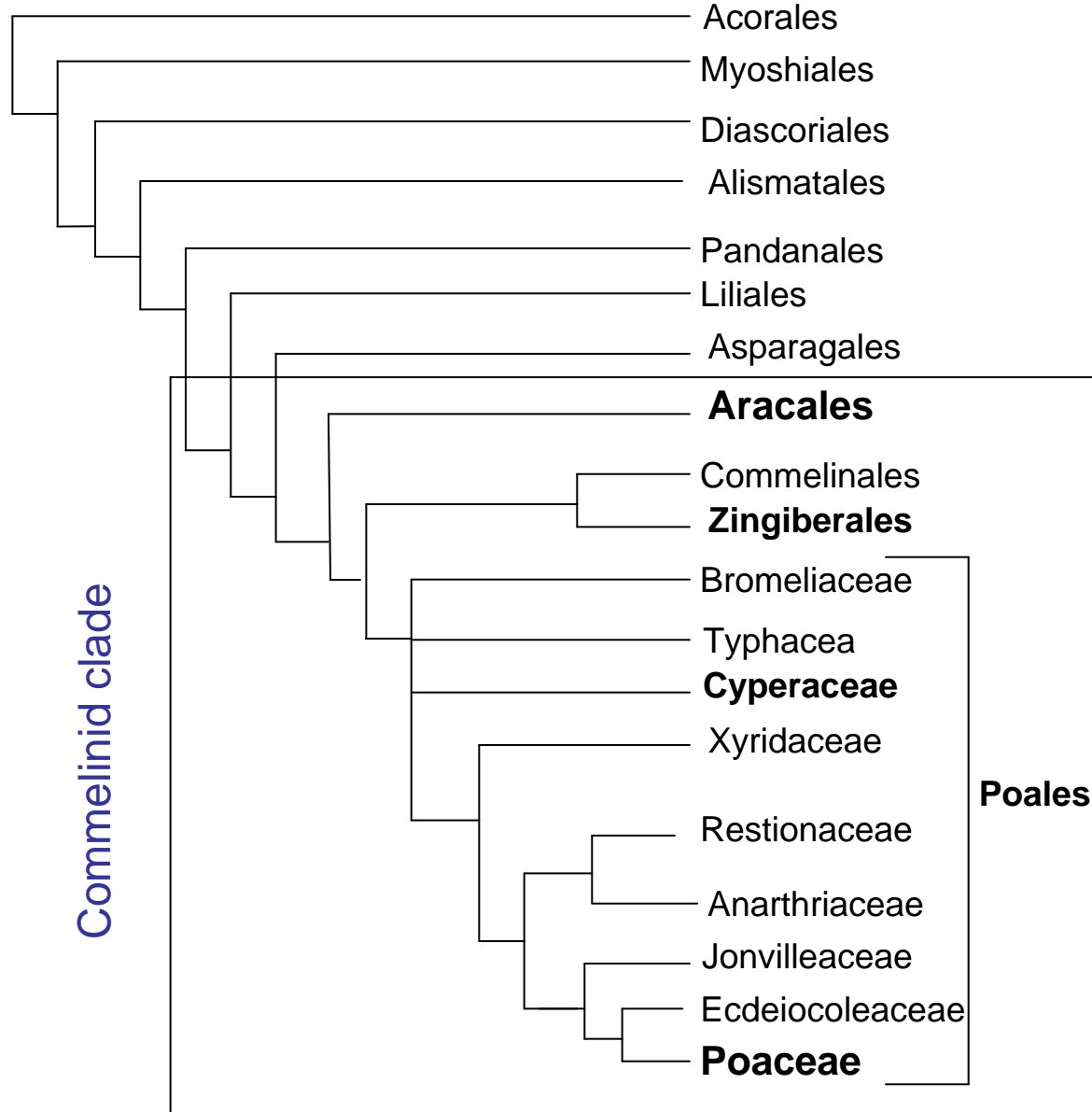
- Synergy from the multi-species approach (models!)
- Empowers minor (e.g., orphan) species
- Illuminates significance and novelty by the degree of gene conservation

## ★ How is it done?

- Comparisons of sequences (genomic, expressed)
- Comparisons of genetic maps
- Comparisons of relationships between gene characteristics and organismal biology



# A Monocot Phylogenetic Tree



# Family

# Poaceae (= Grasses)

## Subfamily

Pooideae Erhartoideae Chloridoideae Panicoideae

## Supertribe

Triticodae Poodae Oryzodae Panicodae Andropogonodae

## Tribe

Triticeae Aveneae Poeae Oryzeae Chlorideae Paniceae Andropogoneae

## Genus

Brachypodium  
Triticum (Wheat)  
Secale (Rye)  
Hordeum (Barley)  
Avena (Oats)  
Festuca  
Lolium (Ryegrass)  
Oryza (Rice)  
Eragrostis (Tef millet)  
Eleusine (Finger millet)  
Pennisetum (Pearl millet)  
Setaria (Foxtail millet)  
Zea (Maize)  
Sorghum  
Saccharum (Sugarcane)



# Oryzoideae



Rice



Sugar cane



Sorghum



Pearl millet

# Panicoideae

Foxtail millet



Maize

# Pooideae



Wheat



# Grass Evolutionary Tree

Family

Poaceae

Subfamily

Pooideae Bambusoideae Chloridoideae Panicoideae

Supertribe

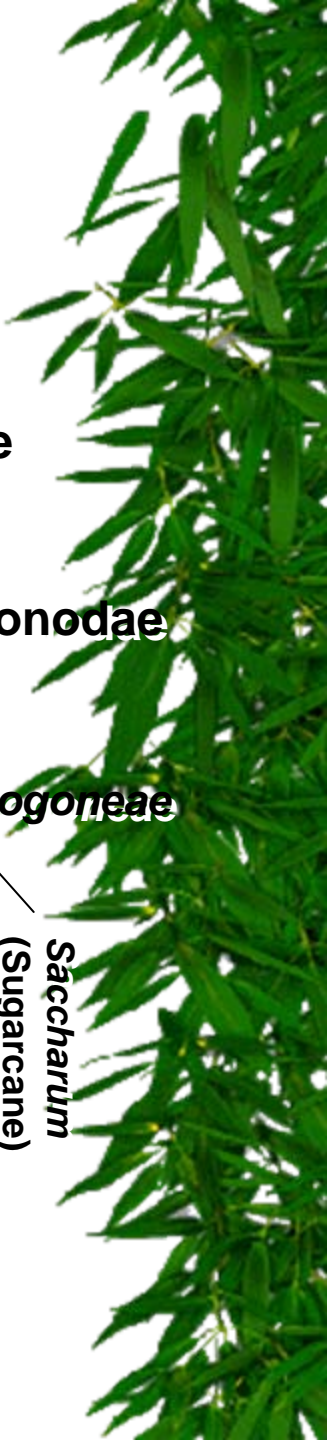
Triticodae Poodae Oryzodae Panicodae Andropogonodae

Tribe

Triticeae Aveneae Poeae Oryzeae Chlorideae Paniceae Maydeae Andropogoneae

Genus

Brachypodium  
 Triticeae  
 Triticum (Wheat)  
 Secale (Rye)  
 Hordeum (Barley)  
 Aveneae  
 Avena (Oats)  
 Poeae  
 Festuca  
 Lolium  
 (Ryegrass)  
 Oryzodae  
 Oryza (Rice)  
 (tef)  
 Eragrostis  
 Chlorideae  
 Eleusine  
 (Finger millet)  
 Panicoideae  
 Paniceae  
 Pennisetum  
 (Pearl millet)  
 Setaria  
 (Foxtail millet)  
 Maydeae  
 Coix  
 Tripsacum  
 Zea (Maize)  
 Andropogonodae  
 Andropogoneae  
 Sorghum  
 (Sugarcane)  
 Saccharum



# *Eragrostis tef*

- \* Allotetraploid ( $2n=4x=40$ ); sexually propagated; self-pollinated; genome size ~730Mb
- \* Domesticated in Ethiopia about 6000 years ago
- \* At last count, constituted 31% of the total acreage of cereals grown in Ethiopia (>2.1 million acres)



# *Eragrostis tef* and lodging

- \* Severely limited by its susceptibility to lodging:
  - There are no *tef* varieties that have good lodging resistance
  - Annual yield losses range from a low of 15% to over 50%, with an average around 25%





# *Striga* on pearl millet in Mali



# Oryzoideae



Rice



Sugar cane



Sorghum



Pearl millet

# Panicoideae

# Pooideae



Wheat



Maize

Foxtail millet



# Foxtail millet growing in Northern China



# The Foxtail Millet As a Model Plant

**A close relative of pearl millet, switchgrass and other Paniceae**

**~ 500 Mb genome, recently sequenced**

**True diploid, easily crossed, small plant size, short generation time**

**Significant research community working on crop improvement and genetic characterization**

**Genetic map, excellent colinearity with rice**

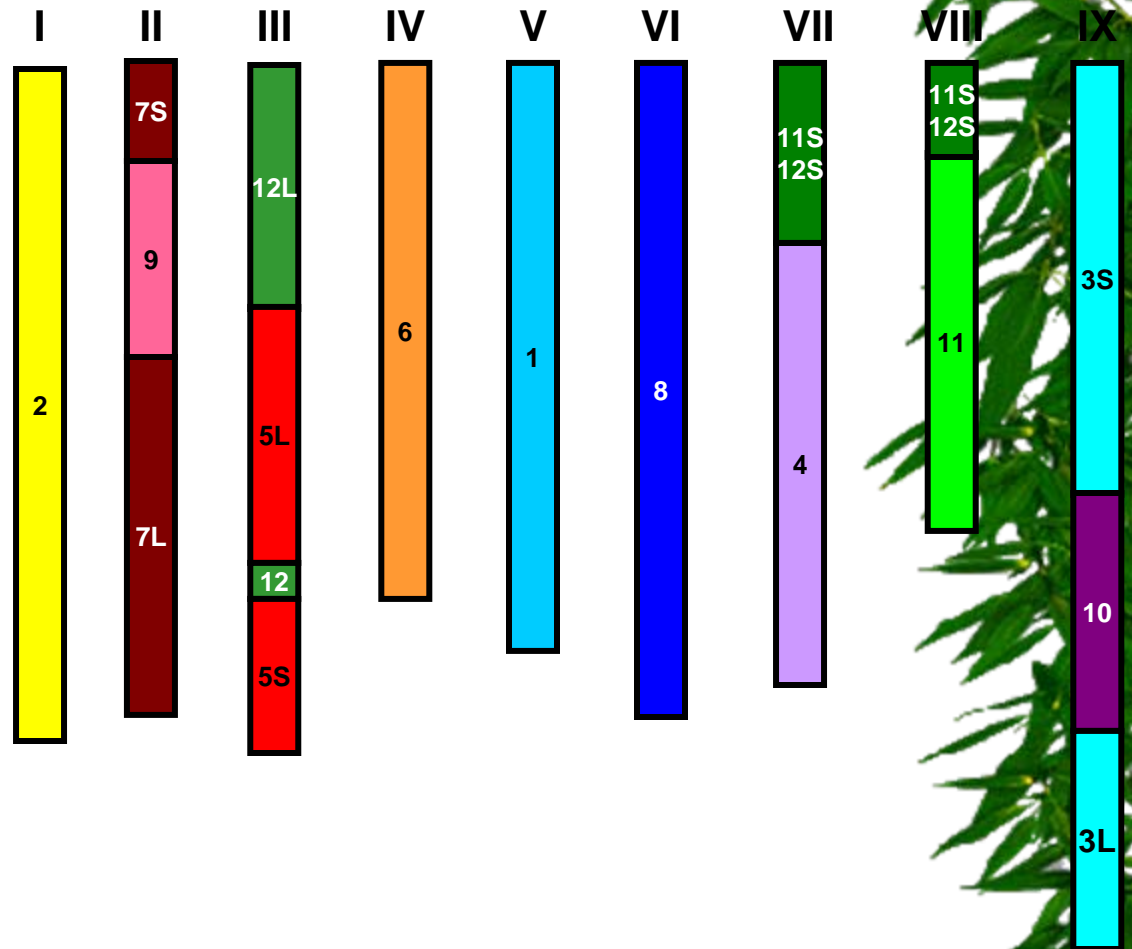
**Fosmid and EST libraries, preliminary microarrays**

**Transformation**





# Foxtail millet (n=9) – rice (n=12)



# Genetic maps and their use



## Genetic maps and their use

**Can identify common genes involved in common processes,  
across species**



## Genetic maps and their use

**Can identify common genes involved in common processes,  
across species**

### **Caveats**

**Lots of genes in small genomic segments**

**Lots of local gene duplication**



# Genetic maps and their use

**Can identify common genes involved in common processes, across species**

## **Caveats**

**Lots of genes in small genomic segments**

**Lots of local gene duplication**

## **Technological changes**

**Mapping by genomic sequence analysis: quick, cheap, dense**  
**Genetic and bioinformatic tools to differentiate paralogs from orthologs**



# Comparing genetic maps



## Comparing genetic maps

**Maps from closely related species are likely to be the most similar, but there are dramatic exceptions (e.g., rye, wheat, barley; foxtail millet, pearl millet, rice).**



## Comparing genetic maps

**Maps from closely related species are likely to be the most similar, but there are dramatic exceptions (e.g., rye, wheat, barley; foxtail millet, pearl millet, rice).**

**Ploidy issues.**

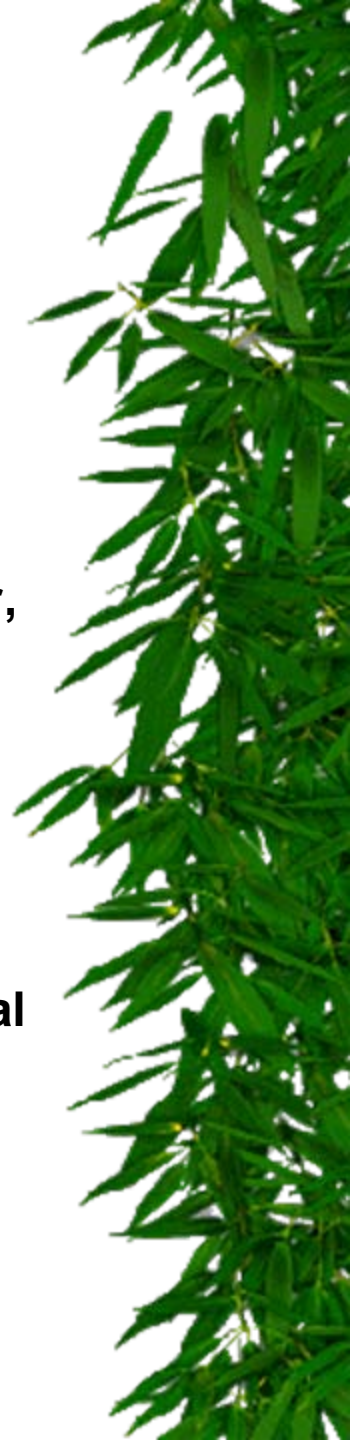


## Comparing genetic maps

**Maps from closely related species are likely to be the most similar, but there are dramatic exceptions (e.g., rye, wheat, barley; foxtail millet, pearl millet, rice).**

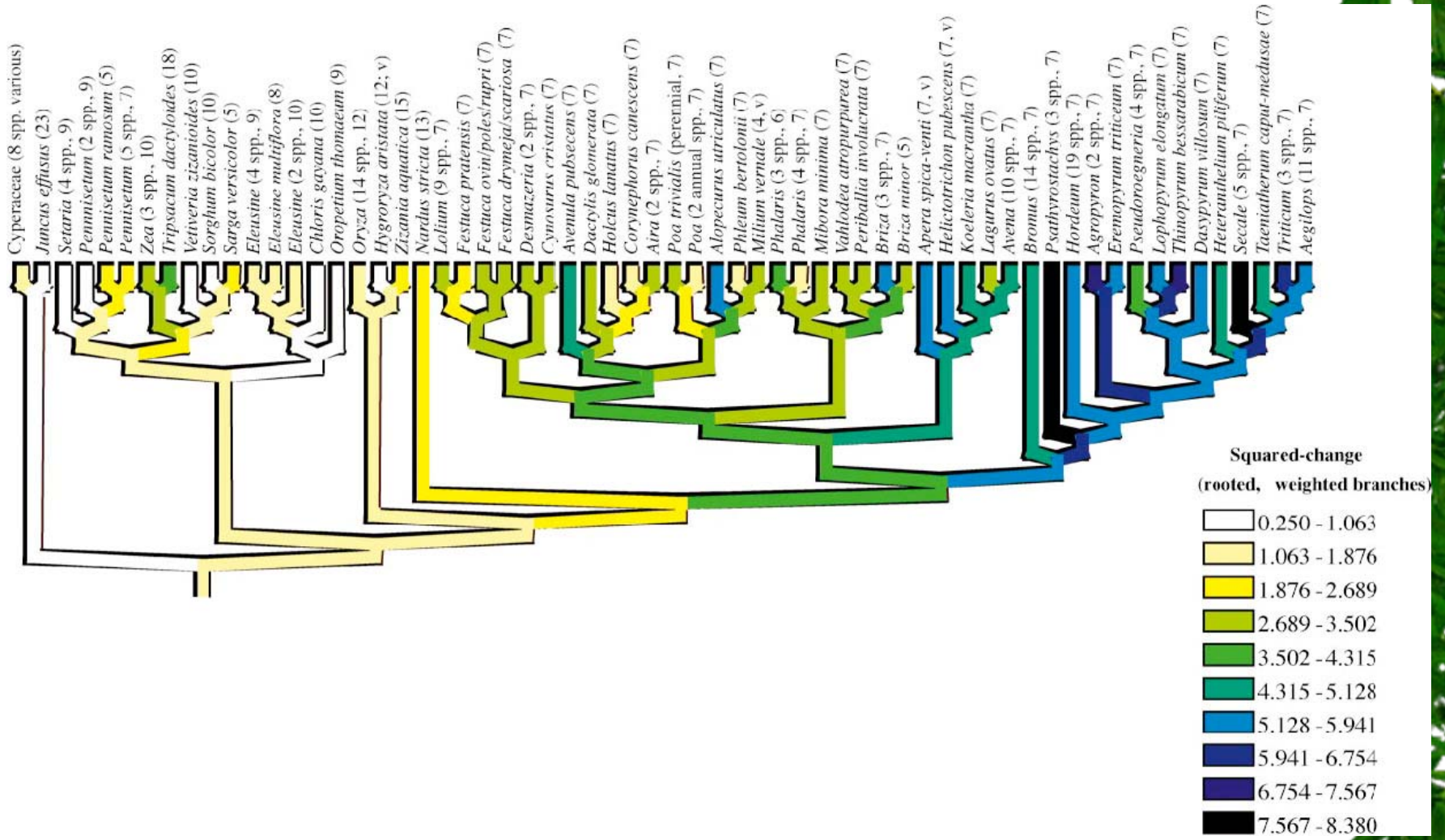
**Ploidy issues.**

**Are the same phenotypes really being compared? (Precise phenotype characterization is now the bottleneck in agricultural genetics and genomics.)**



Plant genomes change very rapidly, by a number of mechanisms, none of which are fully random or quantitatively conserved across species





What is responsible for the variation in plant genome size?

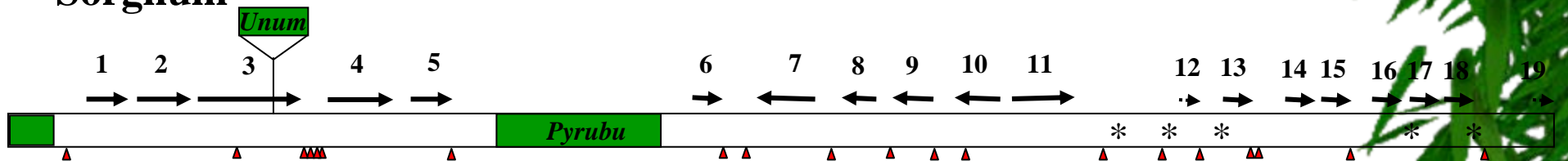


What is responsible for the variation in plant genome size?

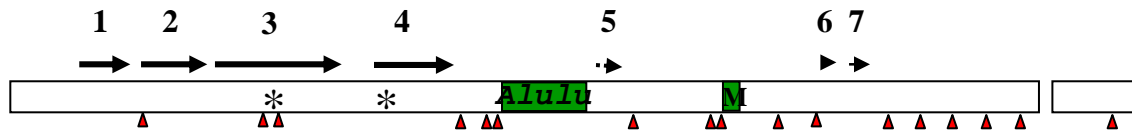
**Different quantities of transposable elements**



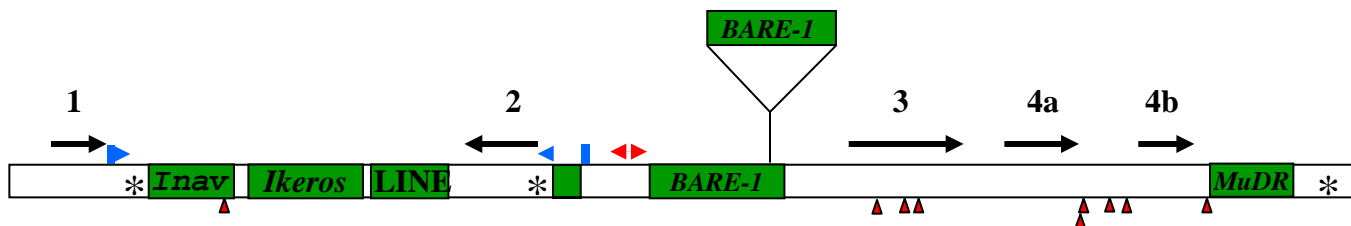
## Sorghum



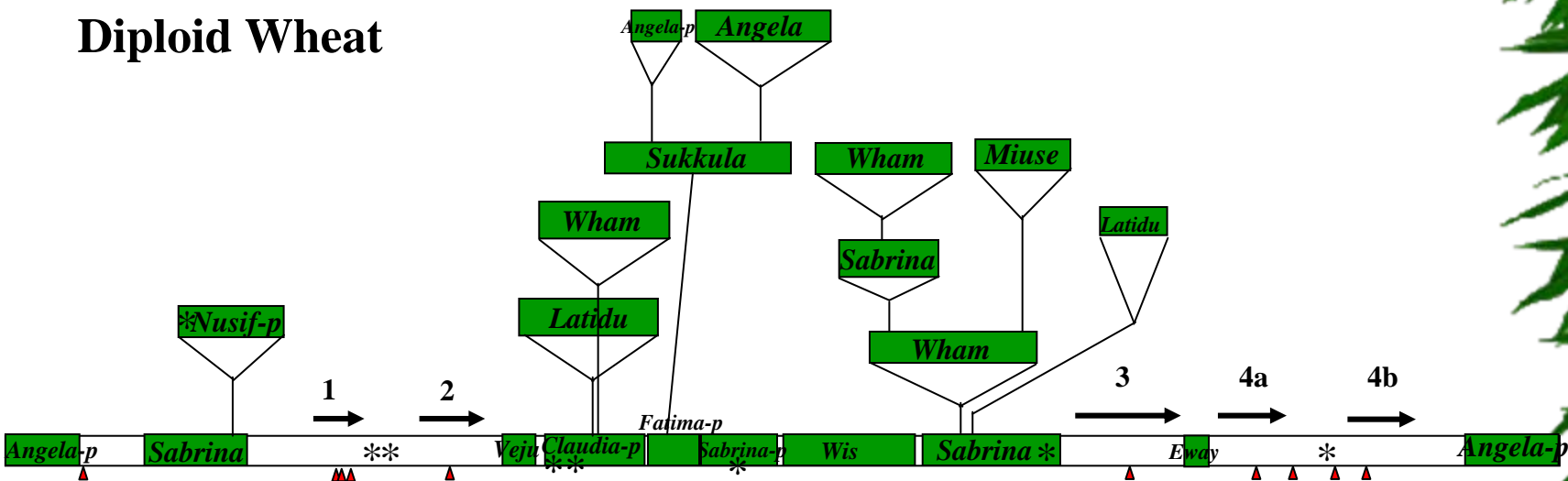
## Rice



## Barley



## Diploid Wheat



What is responsible for the variation in plant genome size?

**Different quantities of transposable elements**



What is responsible for the variation in plant genome size?

**Different quantities of transposable elements**

**Different histories of polyploidy**



What is responsible for the variation in plant genome size?

**Different quantities of transposable elements**

**Different histories of polyploidy**

**Different rates of DNA removal**



How, and how fast, is DNA removed from an angiosperm nuclear genome?



Most DNA is removed by illegitimate recombination



Most DNA is removed by illegitimate recombination

**Non-functional (i.e., unselected) DNA is rapidly lost from plant nuclear genomes, routinely with half lives of less than 2 million years**



Most DNA is removed by illegitimate recombination

**Non-functional (i.e., unselected) DNA is rapidly lost from plant nuclear genomes, routinely with half lives of less than 2 million years**

**An example: more than 300 Mb removed from the rice genome in less than 2 million years**



# The Search for a Mechanism for DNA Removal



I-Scel creation of double-strand  
breaks in maize as an assay for  
inaccurate DNA repair



I-Scel creation of double-strand  
breaks in maize as an assay for  
inaccurate DNA repair

**Engineer I-Scel (18 bp) site into one plant**



I-Scel creation of double-strand breaks in maize as an assay for inaccurate DNA repair

**Engineer I-Scel (18 bp) site into one plant**  
**Engineer I-Scel enzyme expression into a second plant**



I-Scel creation of double-strand breaks in maize as an assay for inaccurate DNA repair

**Engineer I-Scel (18 bp) site into one plant**

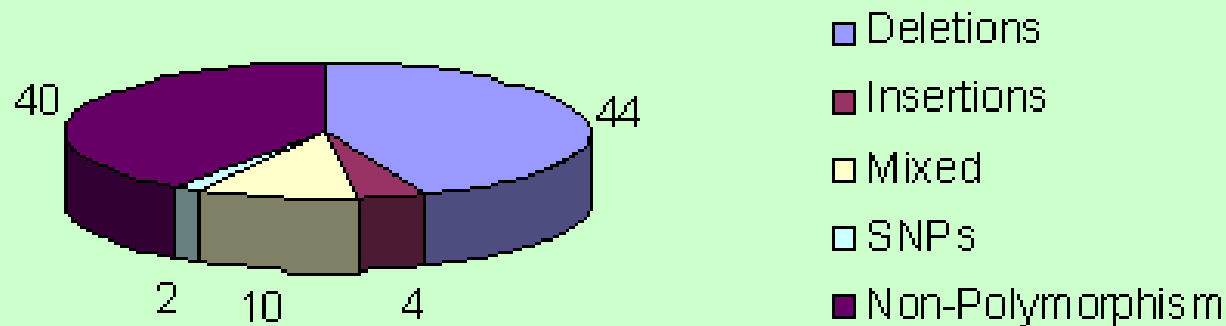
**Engineer I-Scel enzyme expression into a second plant**

**Cross these two plants, score progeny for I-Scel site status**



# I-SceI creation of double-strand breaks in maize as an assay for inaccurate DNA repair

## Classification of DSB Repair Sequences





What is the structure of the average  
flowering plant genome?

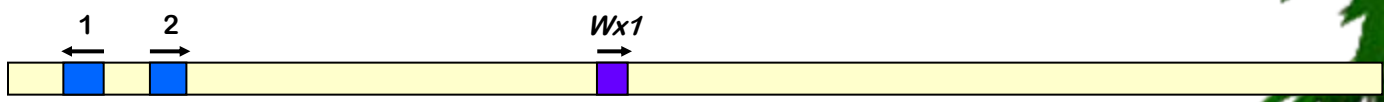


Islands of genes in seas of repetitive DNA.

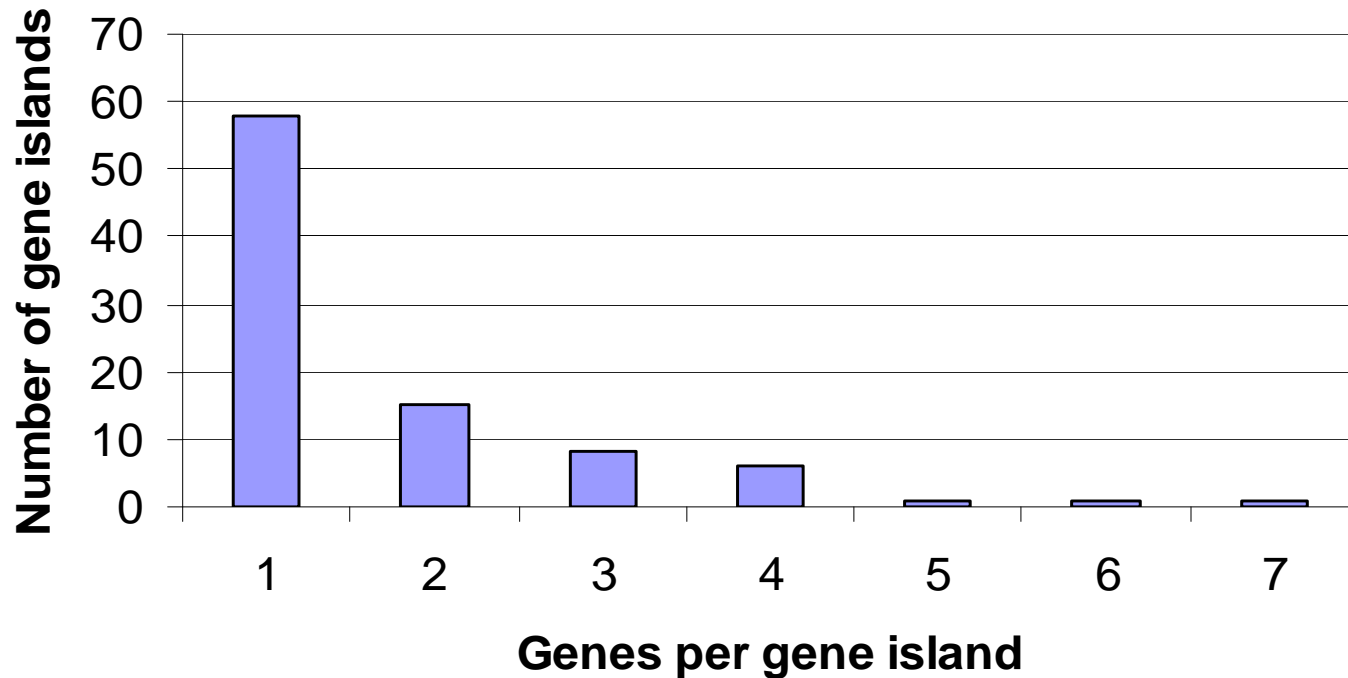




Maize: 143 kb



Most maize gene islands contain only one verified gene



# Gene Content Variation in the Angiosperms



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**
  - **Truncated genes missed as true homologs or paralogs**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**
  - **Truncated genes missed as true homologs or paralogs**
- \* **Genes are lost**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**
  - **Truncated genes missed as true homologs or paralogs**
- \* **Genes are lost: the interesting case of TIR-NBS-LRR genes**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**
  - **Truncated genes missed as true homologs or paralogs**
- \* **Genes are lost**
- \* **Genes do amplify and diverge**



# Gene Content Variation in the Angiosperms

- \* **“Fully Sequenced” genomes of the same ploidy have been reported to vary as much as 2X in gene content**
- \* **Most of this variation is caused by artifacts in annotation**
  - **Low copy # TEs annotated as genes**
  - **Gene fragments within TEs annotated as genes**
  - **Truncated genes missed as true homologs or paralogs**
- \* **Genes are lost**
- \* **Genes do amplify and diverge**
- \* **Genes do move**



More common than presence/absence: divergence (the example of maize *Tb1*)

**Teosinte**

**WT**

***tb1***



## Conclusions – Overall Genome Structure

**Plant genomes add DNA quickly by polyploidy, segmental duplication and transposon amplification**



## Conclusions – Overall Genome Structure

**Plant genomes add DNA quickly by polyploidy, segmental duplication and transposon amplification**

**Plant genomes remove DNA quickly, mostly through illegitimate and unequal recombination**



## Conclusions – Overall Genome Structure

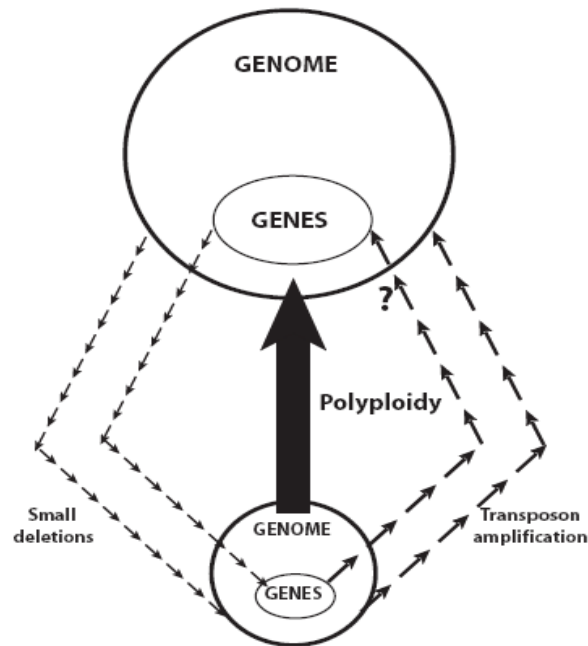
**Plant genomes add DNA quickly by polyploidy, segmental duplication and transposon amplification**

**Plant genomes remove DNA quickly, mostly through illegitimate and unequal recombination**

**The relative aggressiveness of these processes can vary dramatically between plant lineages and between chromosomal locations**



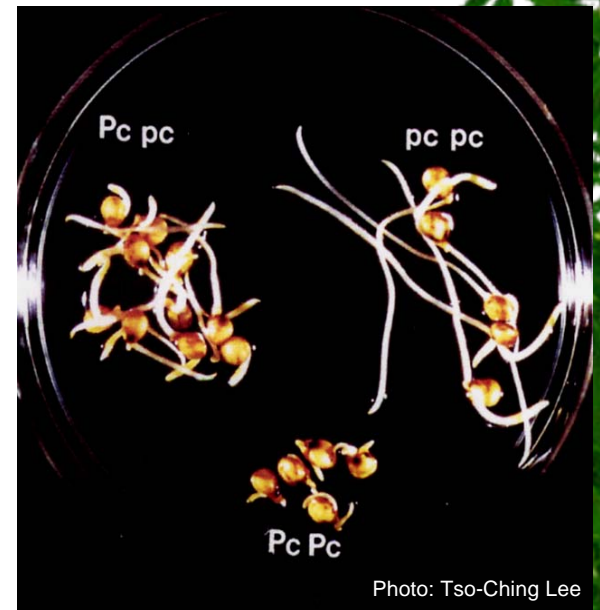
# Genome Dynamics



## Milo disease: root and crown rot in sorghum

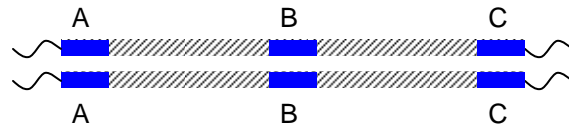
Retarded root development, root and crown rot, red discoloration on the stalk, chlorotic and necrotic spots on the leaves

Caused by a phytotoxin secreted by the soil fungus *Periconia circinata*



Sorghum susceptibility is controlled by a semi-dominant locus (*Pc*) that mutates from *Pc* to *pc* at a rate of about 1/4000 gametes

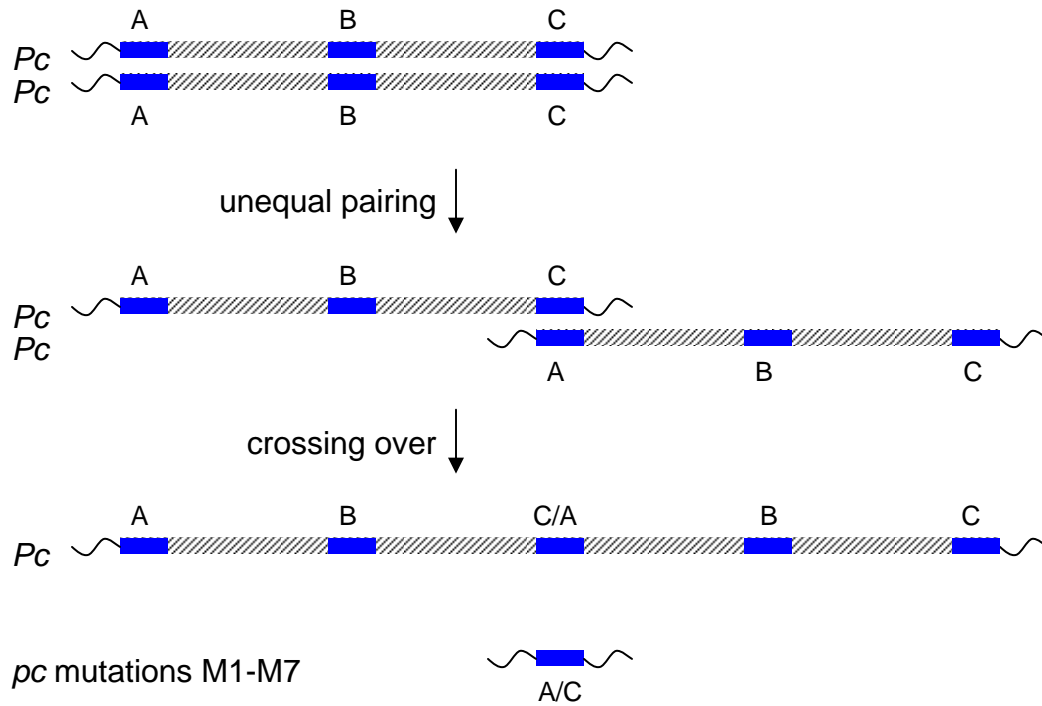
The *Pc* locus is a small gene family of classic disease resistance genes of the CC-NBS-LRR type: corruption of a biotrophic resistance gene by a necrotrophic fungus



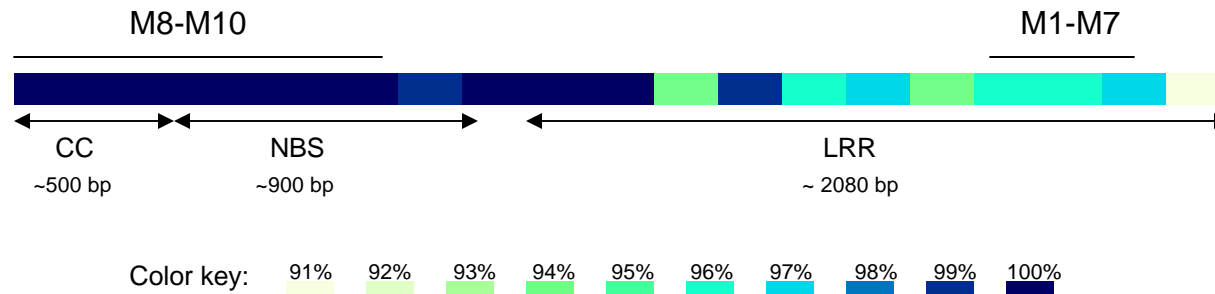
## Twelve $Pc$ to $pc$ mutations

- \* **One is a deletion internal to gene B**
- \* **The other 11 are unequal homologous recombination products that remove gene B**





Unequal recombination at  $Pc$  is primarily located in the LRRs, the least conserved segment of this gene family



# Bennetzen Crew

**Renyi Liu**

**Fang Lu**

**Jianxin Ma**

**Ervin Nagy**

**Ana Pontaroli**

**Clementine Vitte**

# Collaborators

**Katrien Devos – Univ. Georgia**

**Phillip San Miguel – Purdue Univ.**

**Supported by DOE, NSF and USDA  
Competitive Grants**

