



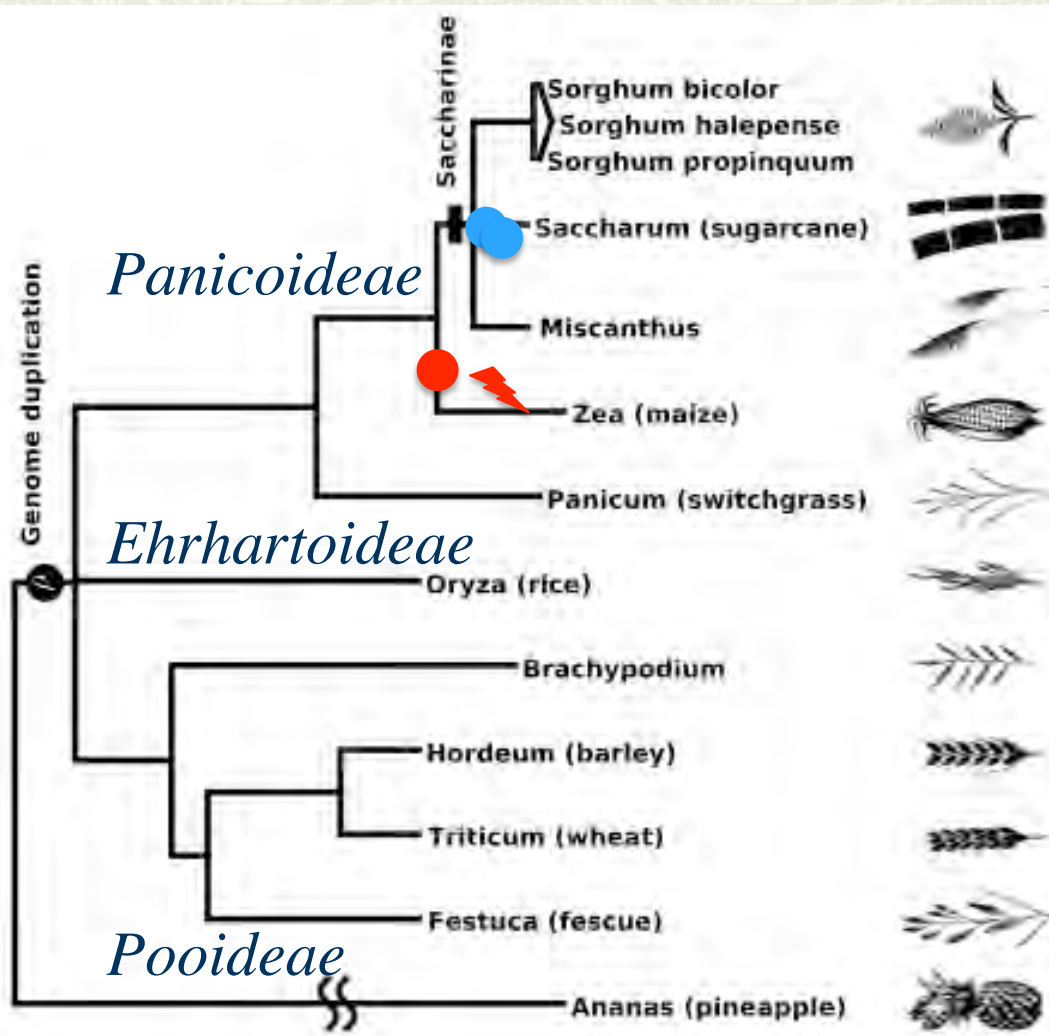
1st National Conference on Biotechnology Lima, Peru, 12 de mayo de 2009

Genomics in the Gramineae and Application to Productivity and Nutritional Quality for Sorghum, Corn, Sugarcane, and Rice



Joachim Messing, Waksman Institute,
Rutgers University, New Jersey, USA

Evolution of grasses



Paterson et al. (2009)
Plant Phys. 149, 125-131

Sequencing entire genomes



Volume 9 Number 12 1981

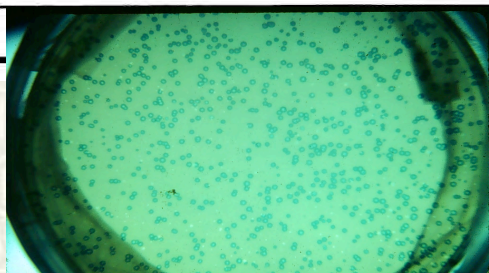
Nucleic Acids Research

The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing

Richard C. Gardner, Alan J. Howarth, Peter Hahn, Marianne Brown-Luedi², Robert J. Shepherd¹ and Joachim Messing^{2,3}

Department of Plant Pathology, and ²Department of Bacteriology, University of California, Davis, CA 95616, USA

Received 21 April 1981



DNA sequencing machines



Volume 9 Number 2 1981

Nucleic Acids Research

A system for shotgun DNA sequencing

Joachim Messing^{††}, Roberto Crea^{*} and Peter H. Seeberg^{**}

[†]Department of Bacteriology, UC Davis, CA 95616, and ^{*}Division of Organic Chemistry and ^{**}Division of Molecular Biology of Genentech, Inc., South San Francisco, CA 94080, USA

Received 30 October 1980

Virus:	3 to 70 genes
Bacterium:	470 to 4,400 genes
Yeast:	6,000 genes
Fruit fly:	14,000 genes
Worm:	19,000 genes
Human:	21,000 genes
Arabidopsis:	23,000 genes
Rice:	28,000 genes
Corn:	40,000 genes

Sorghum Genome Project





Sorghum Genome Sequencing Project

Library Type	Total Reads	HQ Paired HQ	Coverage	Average Insert	Standard Dev.
Small Insert	4,817,407	4,114,610	3.74x	2,443	+/- 385
Medium Insert I	2,661,374	2,374,732	2.32x	6,398	+/- 528
Medium Insert II	2,149,803	1,812,074	1.72x	6,881	+/- 585
Medium Insert III	18,144	16,108	0.01x	8,613	+/- 759
Fosmid	850,443	606,062	0.52x	34,647	+/- 3,812
BAC/SB_BBc	193,920	176,744	0.17x	107,982	+/- 21,811
BAC/SB_BBd	26,112	16,996	0.02x	90,993	+/- 25,014
Total	10,717,203	9,117,326	8.50x		

- Pair wise reads
- Mix of insert sizes
- 735 bp average read

The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers

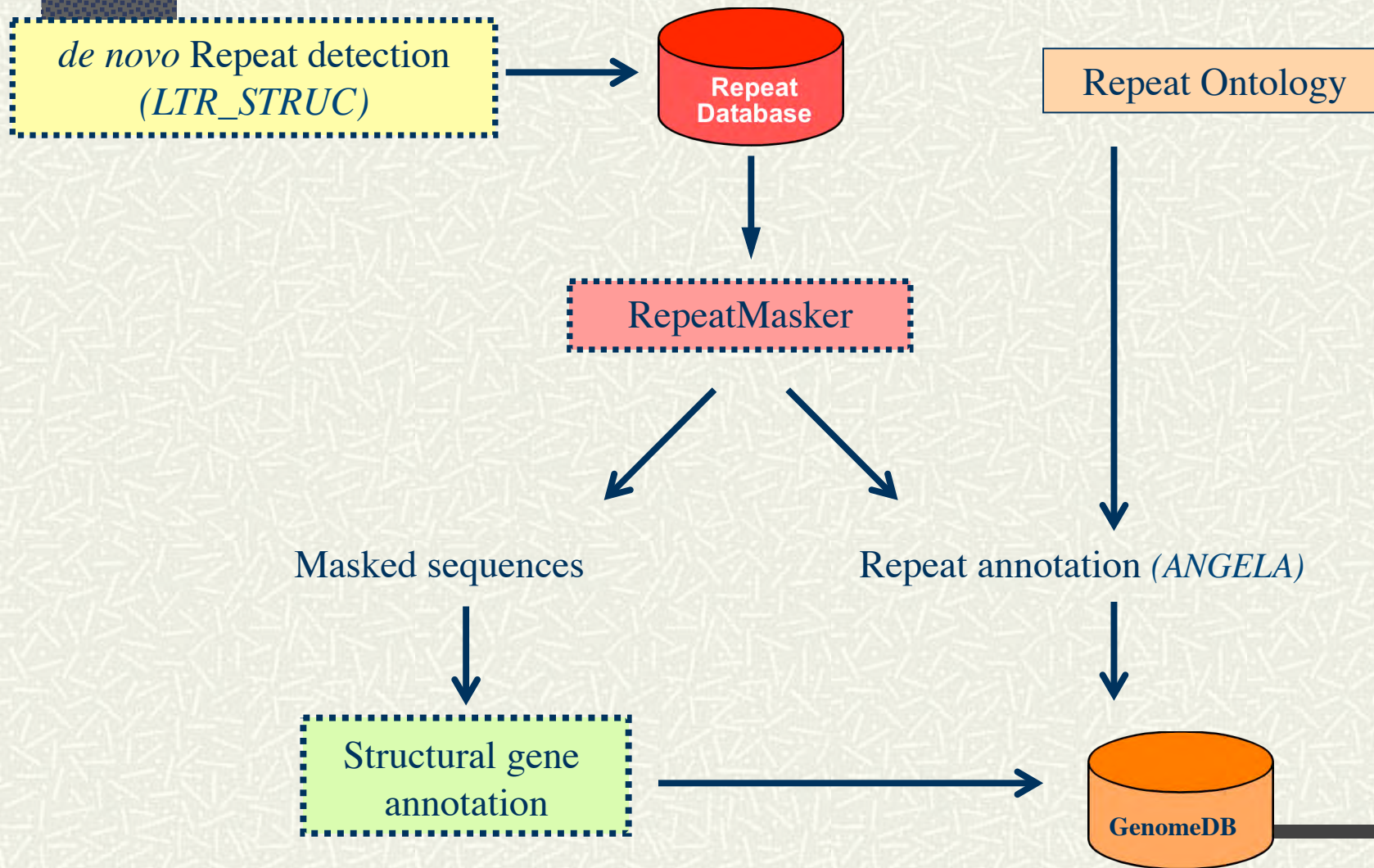
(Recombinant DNA; multiple cloning sites; restriction sites mobilizing element; dideoxy sequencing)

Jeffrey Vieira and Joachim Messing

Department of Biochemistry, University of Minnesota, St. Paul, MN 55108 (U.S.A.)

(Received April 5th, 1982)

Repeat detection and filter

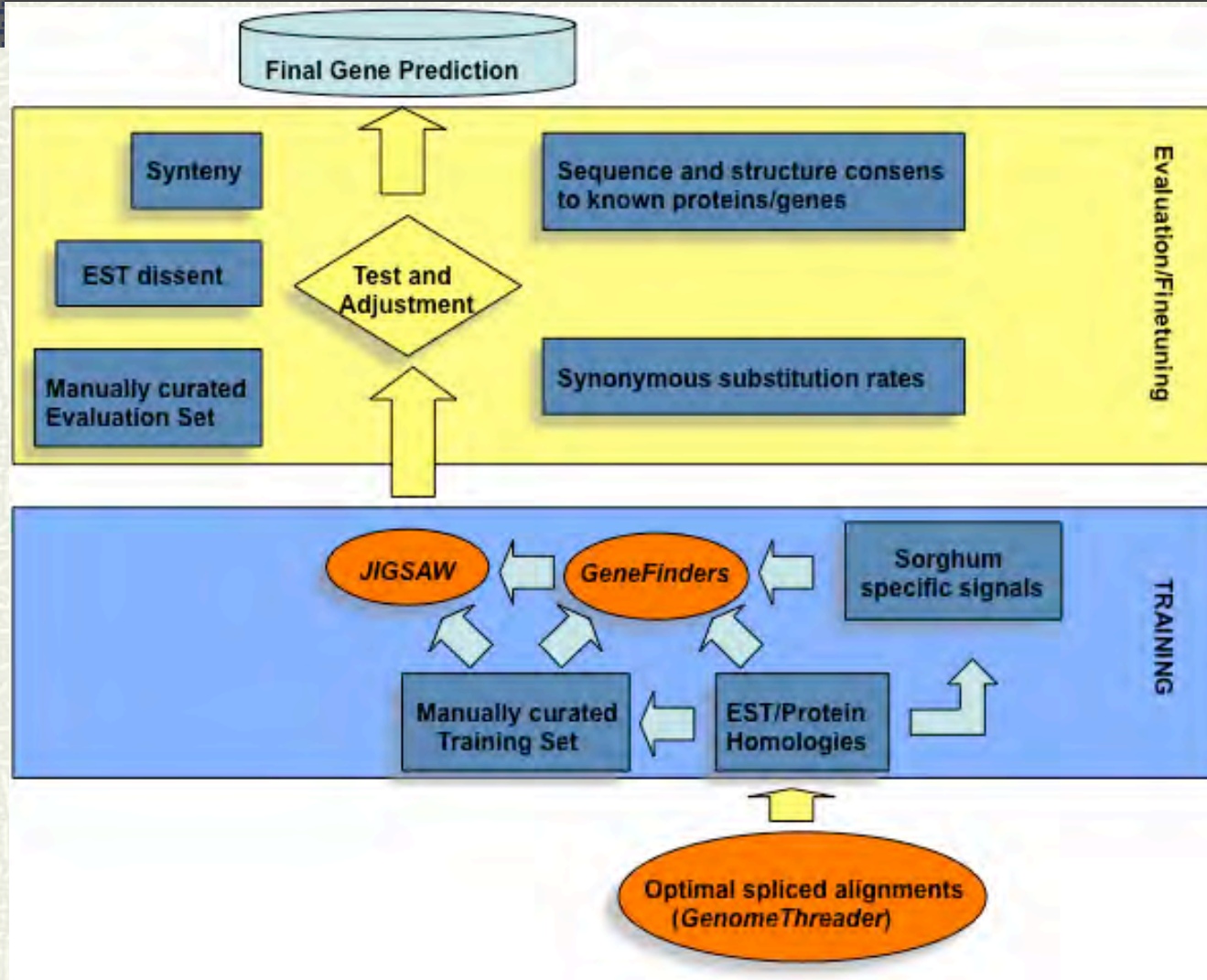


TEs in sorghum, rice, and maize

	SB	OS	ZM*
Class I: Retroelement**	97.88	65.35	98.05
LTR Retrotransposon	97.73	59.49	92.65
Ty1/copia	8.68	6.25	26.84
Ty3/gypsy	31.66	30.49	46.57
unclassified LTR	57.39	22.76	19.24
non-LTR Retrotransposon	0.07	3.15	0.43
LINE	0.07	2.02	0.42
SINE	0.00	1.13	0.01
unclassified retroelement	0.09	2.70	4.97
Class II: DNA Transposon	2.12	34.65	1.95
DNA Transposon Superfamily	0.34	17.83	1.14
CACTA superfamily	0.17	8.68	0.58
hAT superfamily	0.04	1.33	0.12
Mutator superfamily	0.10	4.58	0.18
Tc1/Mariner superfamily	0.00	0.05	0.00
PIF/Harbinger	0.03	0.01	0.09
unclassified super family	0.00	3.18	0.15
MITE	1.48	13.29	0.40
other DNA transposon	0.29	3.52	0.42
Total as percent of genome			
Transposon DNA**	58.2	39.4	80.9
Coding space	13.7	33.0	7.5
Intergenic space incl. regulatory seq.	28.1	27.6	11.6

100 random BACs*
Percent of nucleotides**

Annotation Pipeline



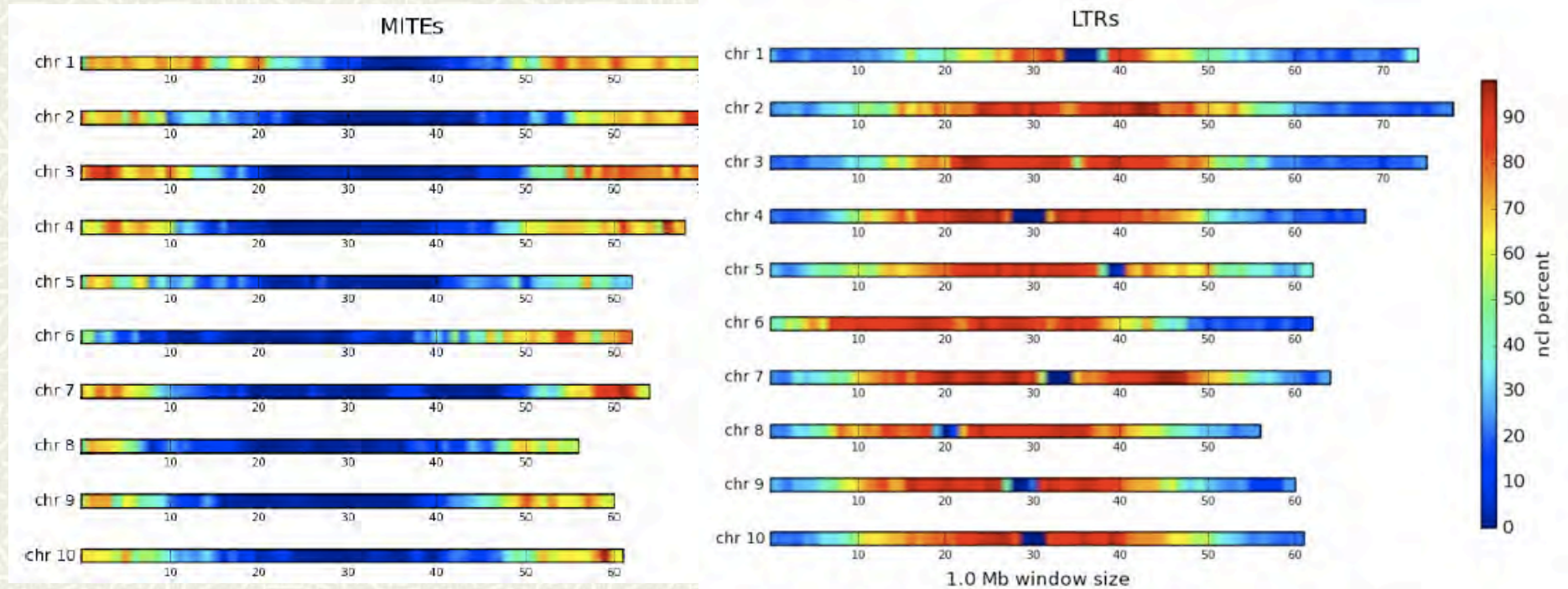


Gene Models in sorghum, rice, and maize

Features	Sorghum	Rice	Maize
	Sb1.4	Rap2	100 Bacs
Genome size	738540932	372089805	14419909
Chromosome assemblies	659229367		
unassembled	79311565		
# genes	27640	28236	330
average # of exons per gene	4.4	4.8	4.6
average exon size [bp]	303	364	234
median exon size [bp]	159	165	128
average intron size [bp]	441	441	545
median intron size [bp]	145	161	177.5
average gene size * [bp]	2616	2468	3040
median gene size * [bp]	1810	1811.5	1940
Average gene density (kb per gene)	26.7	13.2	43.7
GC content [%]			
Overall	46.4	43.6	46.6
Exons	53.5	51.2	55.5
Introns	39.8	37.7	42.5

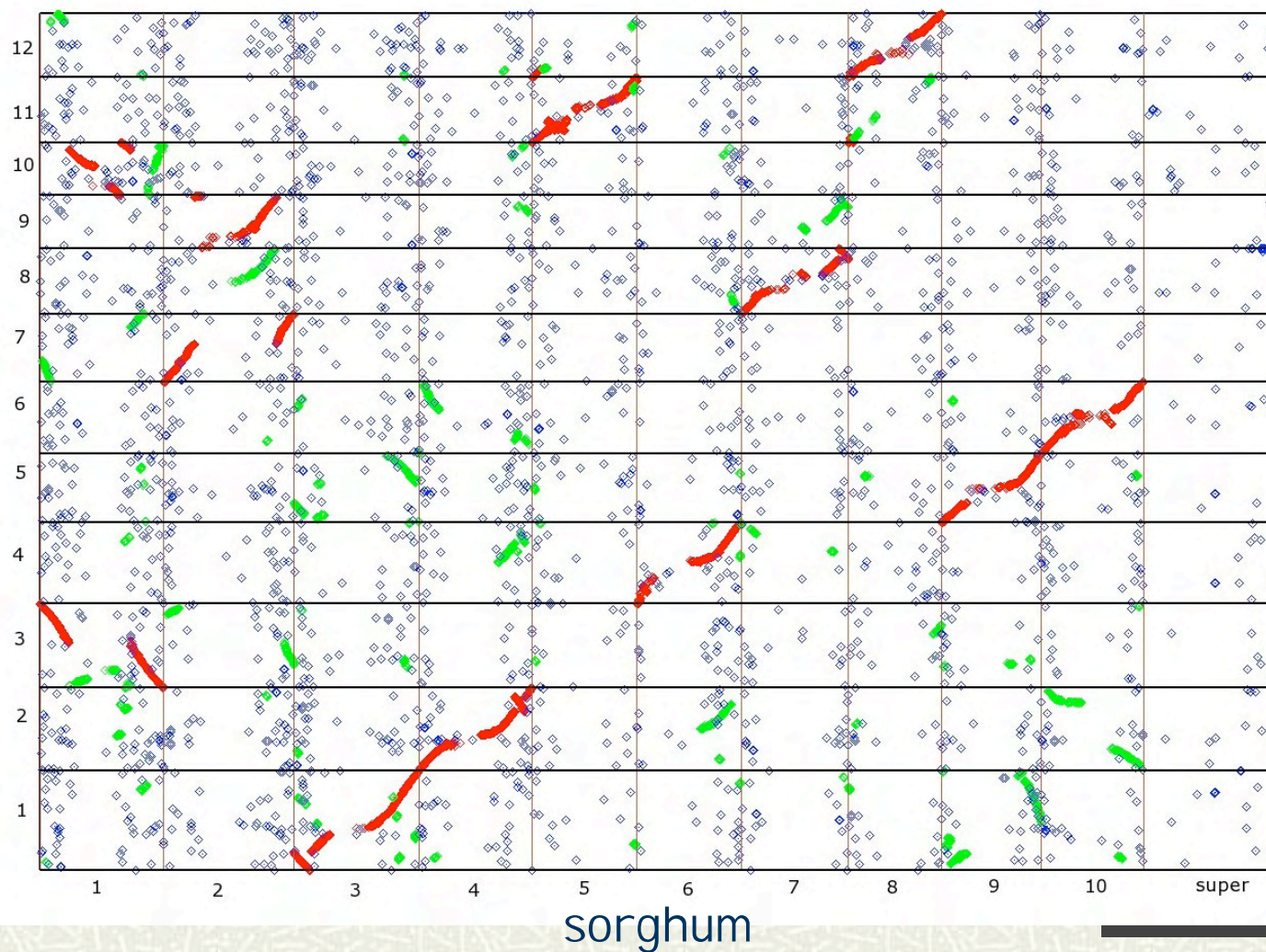
27,640 bona-fide
 5,197 remnants
 727 pseudogenes
 932 TE-related
 34,496 total

Distribution of TEs in sorghum



Synteny between rice and sorghum

rice

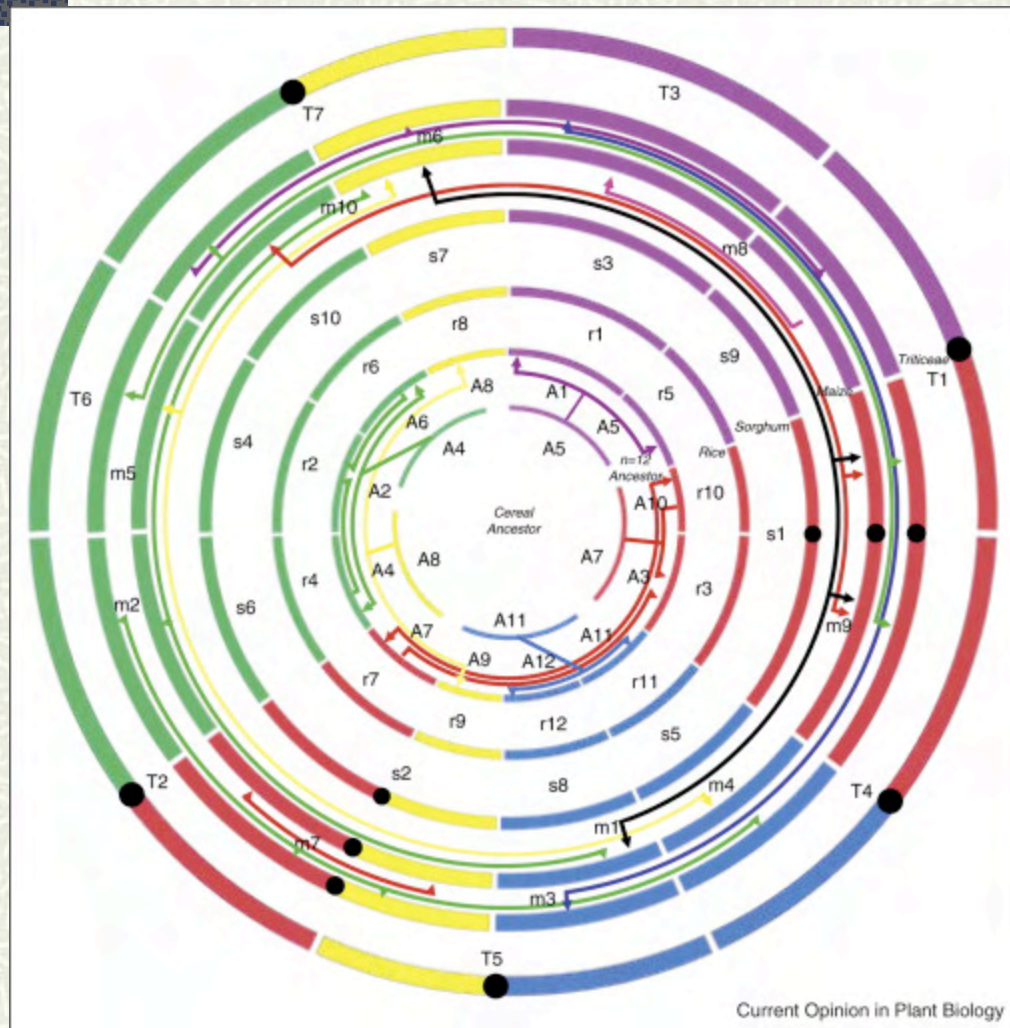


Highlights



- Sorghum and rice genomes are largely collinear with about 28K gene models.
- The sorghum genome expanded by retrotransposition in the pericentromeric regions relative to rice in the last million years.
- Pericentromeric heterochromatin occupies at least 460 mbp (62%) in sorghum versus 63 mbp (15%) in rice.
- Relative higher gene copies (e.g. cytochrome P450, microRNA169) compared to rice may contribute to sorghum's draught tolerance.
- While in rice mutator-like elements (Pack-MULE) create copies of gene (fragments) in sorghum CACTA-like elements and helitrons do.
- Conservation of genes between sorghum and rice by ancestry provides a synergistic reference because of the validation of gene models and their position.

The basic chromosome number in grasses



Based on synteny blocks of multiple pair-wise alignments of orthologous regions, the basic ancestral chromosome number could have been 5.

S. Bolot et al. *Curr Opin Plant Biol*, Dec 2008.



Conclusions from Venn Diagrams

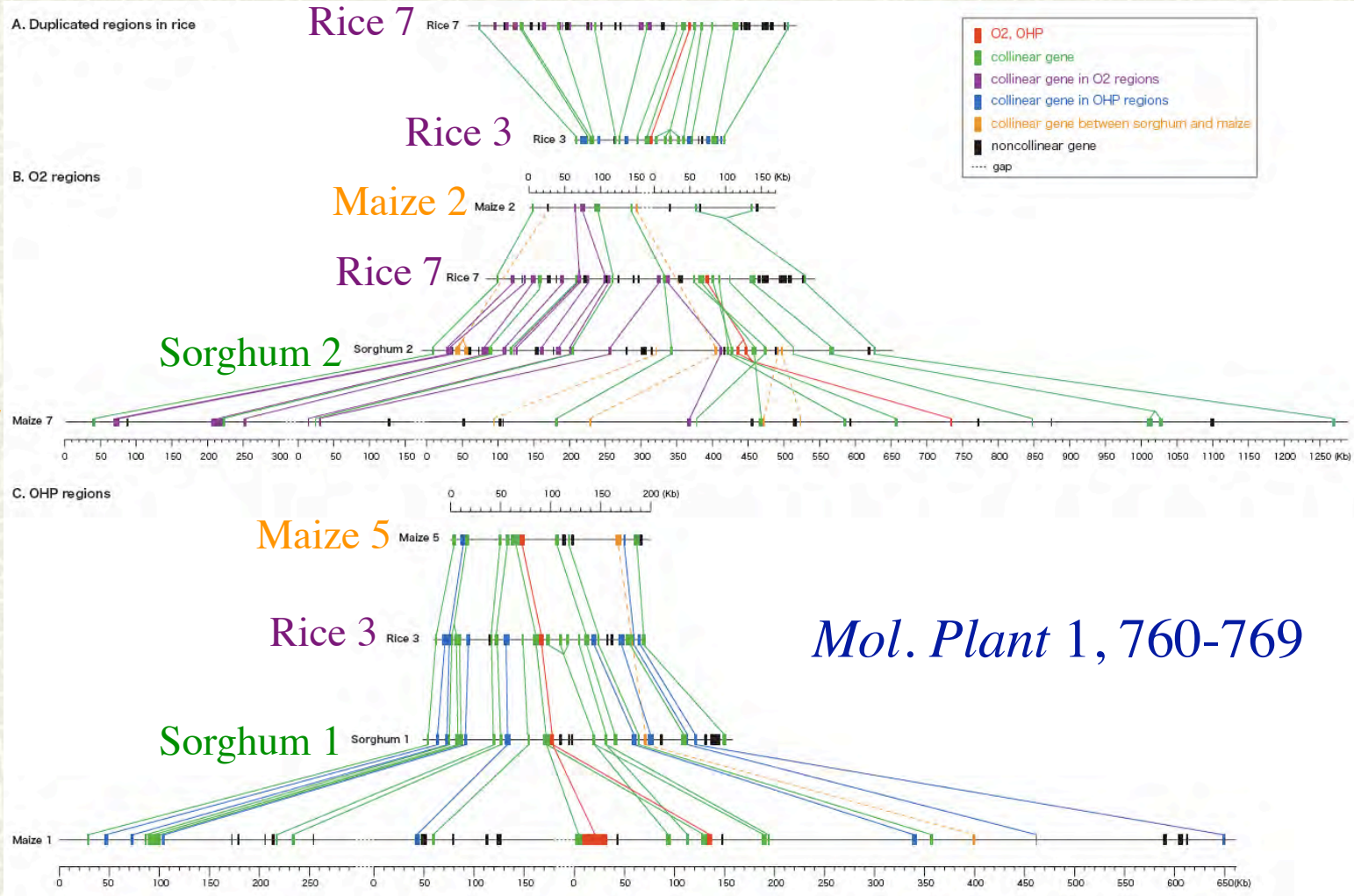
A total of 9,503 (58%) sorghum gene families shared with *Arabidopsis*, rice and poplar: 15,225 (93%) overlapped with at least one other species.

Nearly 94% of high confidence sorghum genes (25,875/27,640) have orthologs in rice, *Arabidopsis*, and/or poplar, and together these gene models define 11,502 ancestral angiosperm gene families.

3,983 (24%) gene families have members only in the grasses sorghum and rice; 1,153 (7%) appear unique to sorghum, which is similar to *Arabidopsis* (6.7%), but fewer in rice (3.6%), and more in poplar (15.7%).

About 2,054 orthologs of our sorghum and TIGR5 rice models are absent from RAP2; ~12,000 TIGR5 models may be TEs or pseudogenes.

Ancient segmental duplication



O2

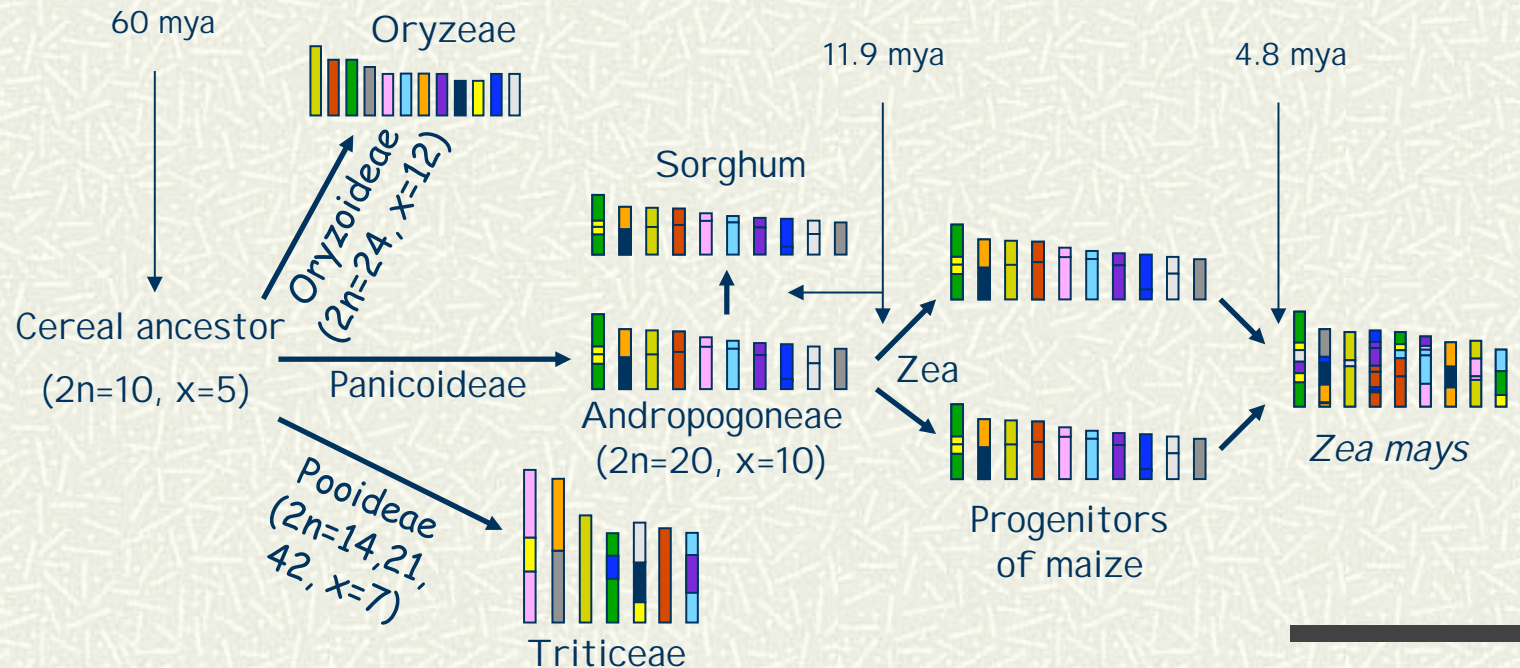
Maize 7

OHP

Maize 1

Chromosome number of the progenitors of maize

Wei et al. (2007) PLoS Genetics 3, 3123

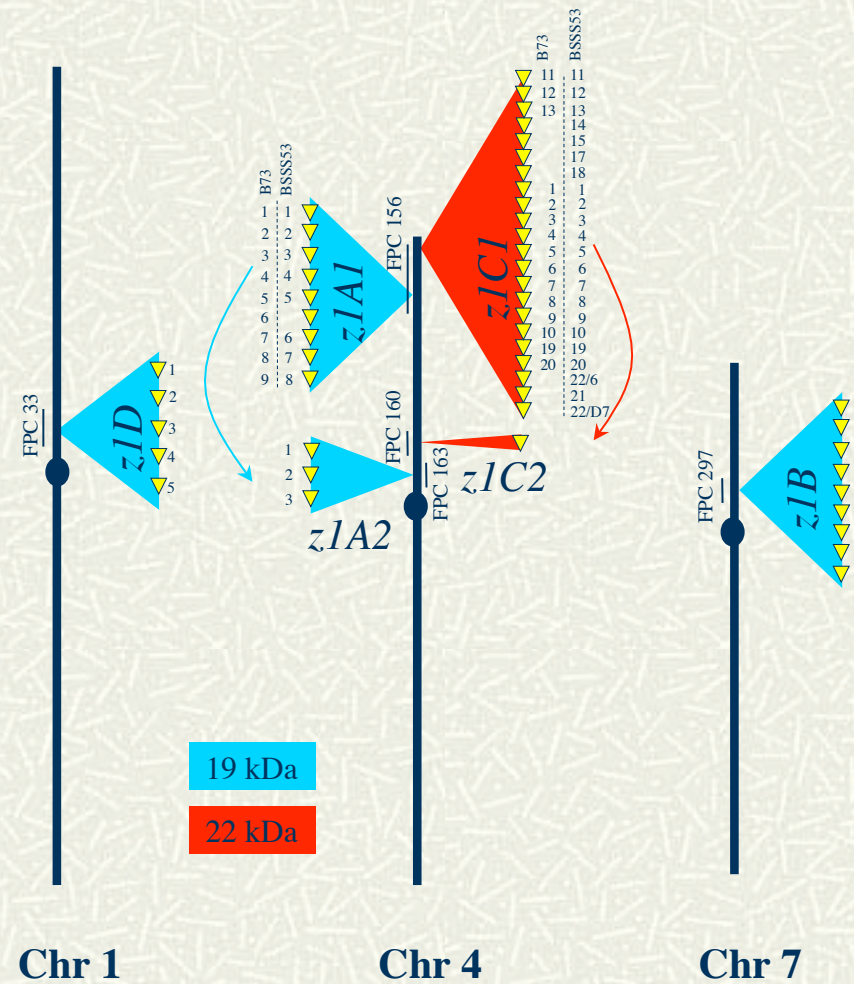




Early duplication in the grass family.

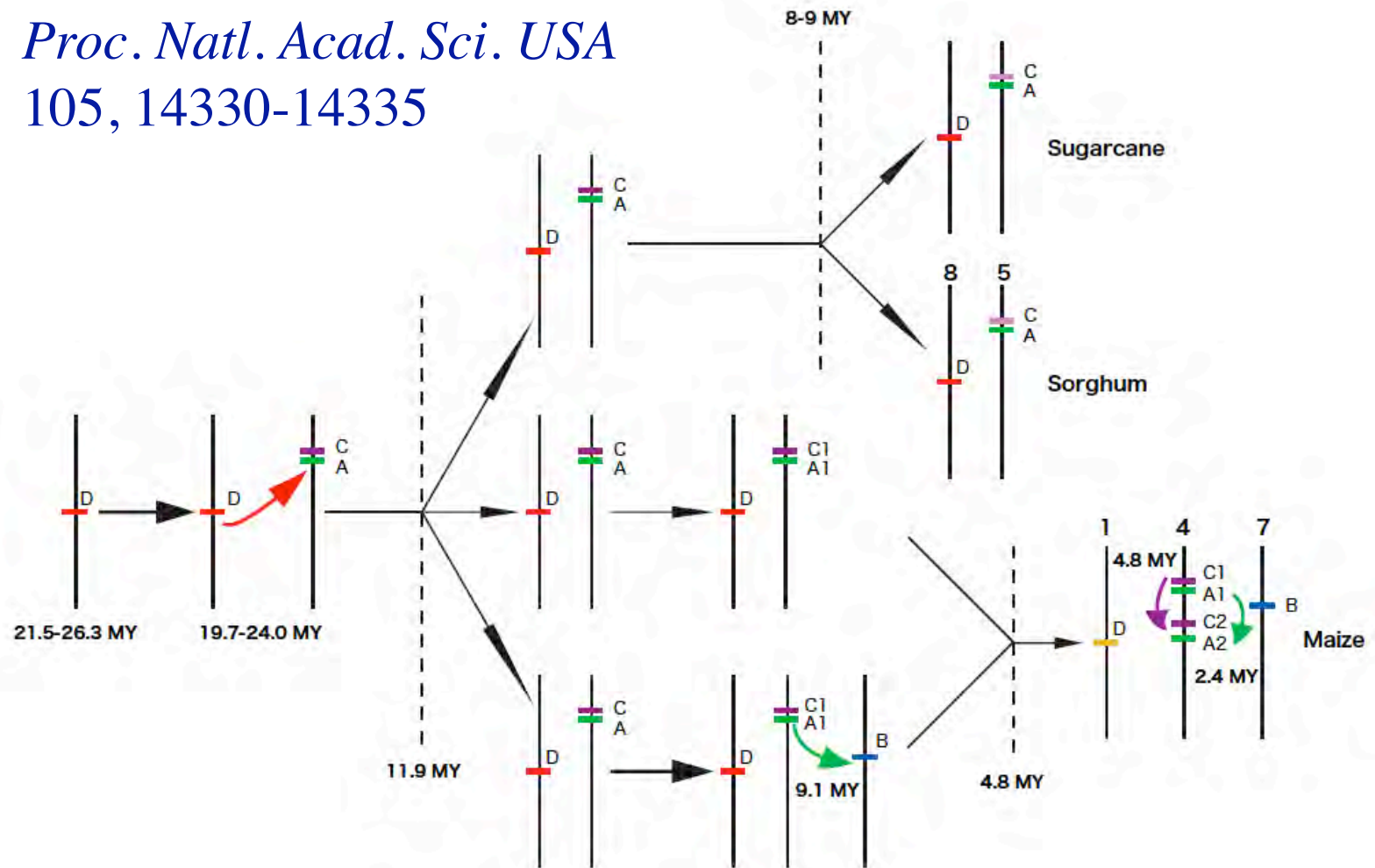
- Segmental duplication about 57 mya in progenitor of rice and maize (50 mya).
- Non collinear genes already before rice and maize split.
- Unequal expansion and divergence by insertions and subfunctionalization.
- Changes are more drastic in maize compared to rice and sorghum.
- Comprise deletions, duplications, conversions, inversions, and translocations.

Alpha prolamins in maize, an example of gene amplification, nutritional quality and heterosis

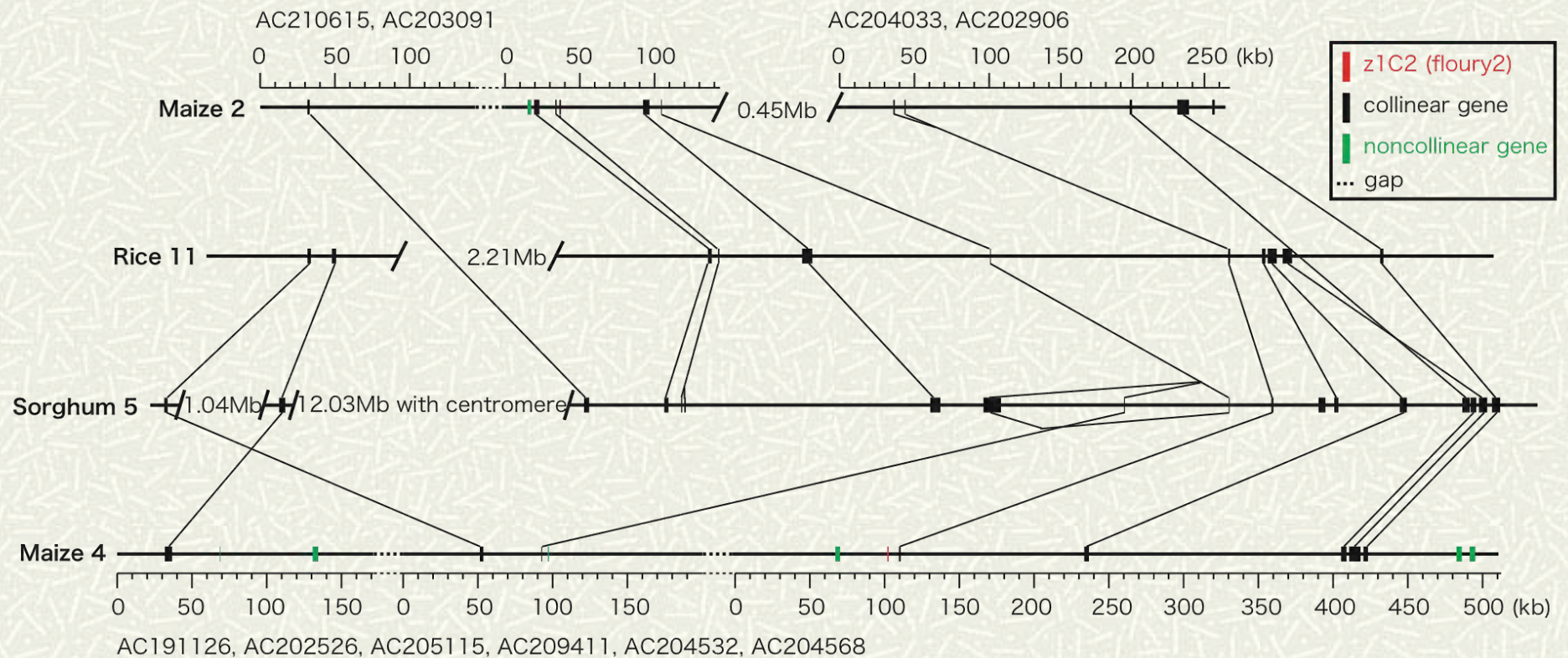


Loci formation during the evolution of the Panacoideae

Proc. Natl. Acad. Sci. USA
105, 14330-14335



Recent insertions of paralogous sequences seem to prefer fragile sites



Proc. Natl. Acad. Sci. USA
105, 14330-14335

Transcriptional regulation of zein genes by O2



basic motif

leucine zipper

MW: 47 kDa (437 a.a) I.P: 4.8

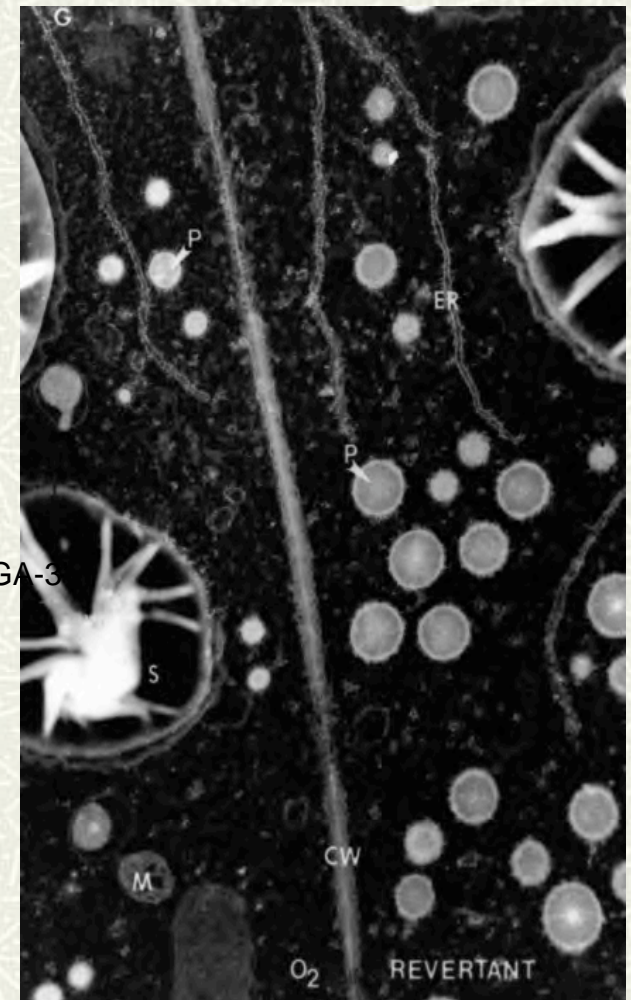
Prolamine Box

O2 Target Sequence

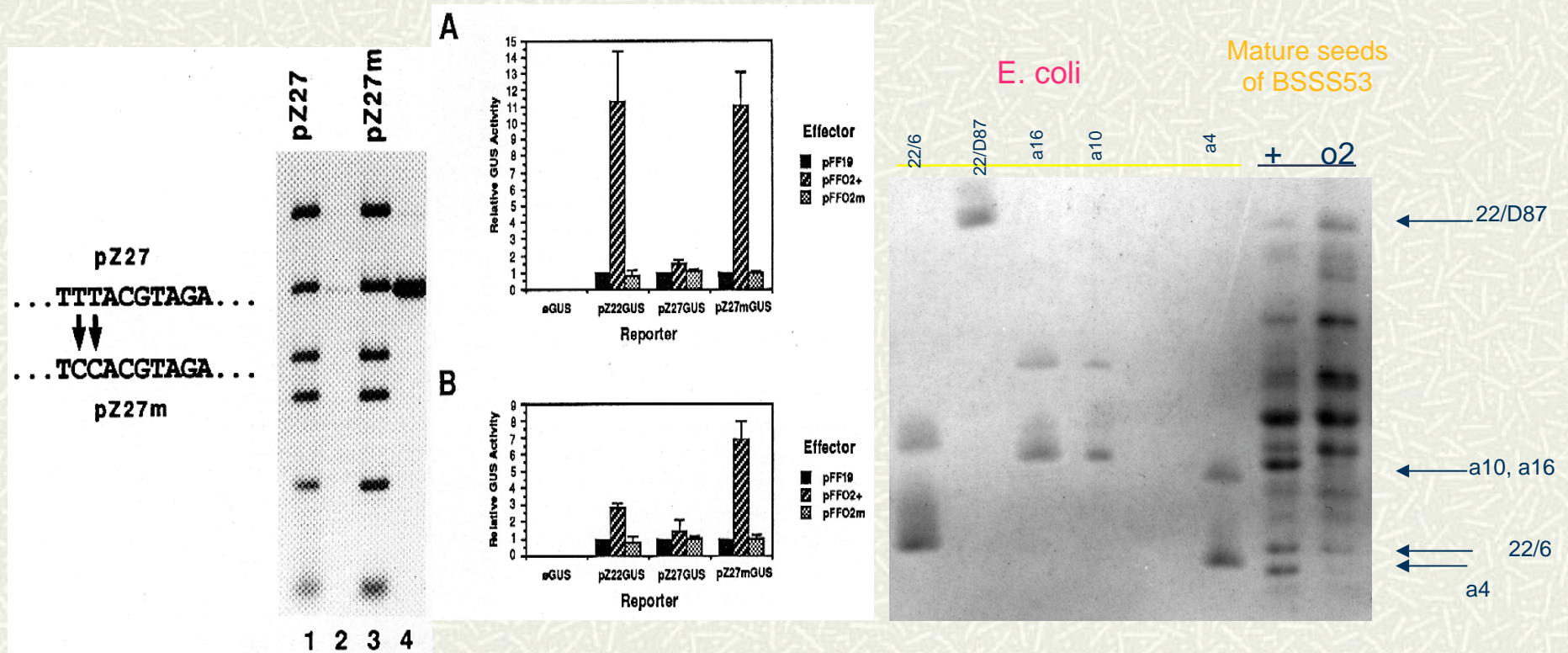
5' -AGGTTGTCACATGTGTAAAGGTGAAGAGATCATGCATGTCATTCCACGTAGATGA-3'

251 nt

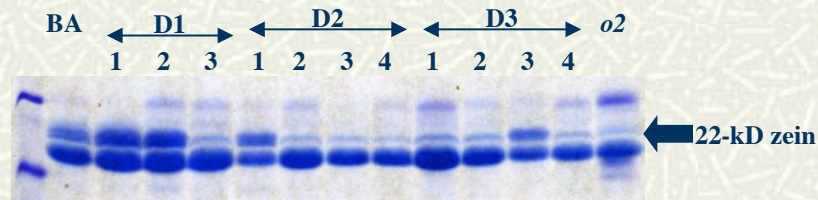
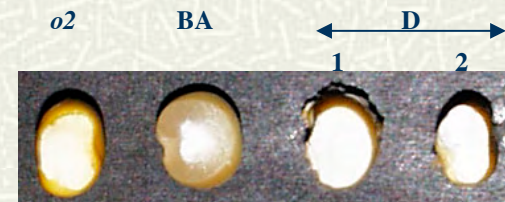
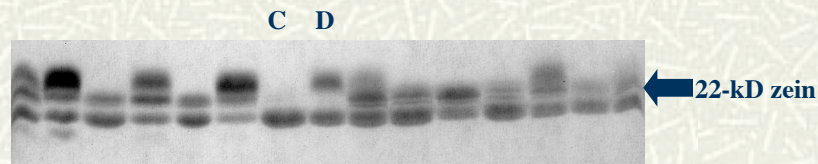
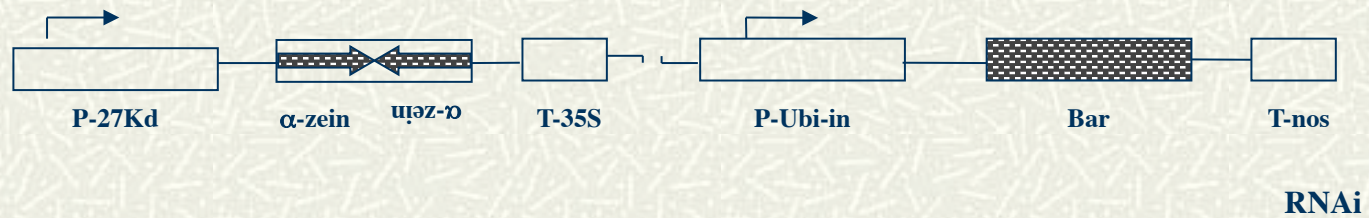
Opaque2 is a bZIP transcriptional factor which specifically recognizes 5'-TCCACGTAGA-3' motif.



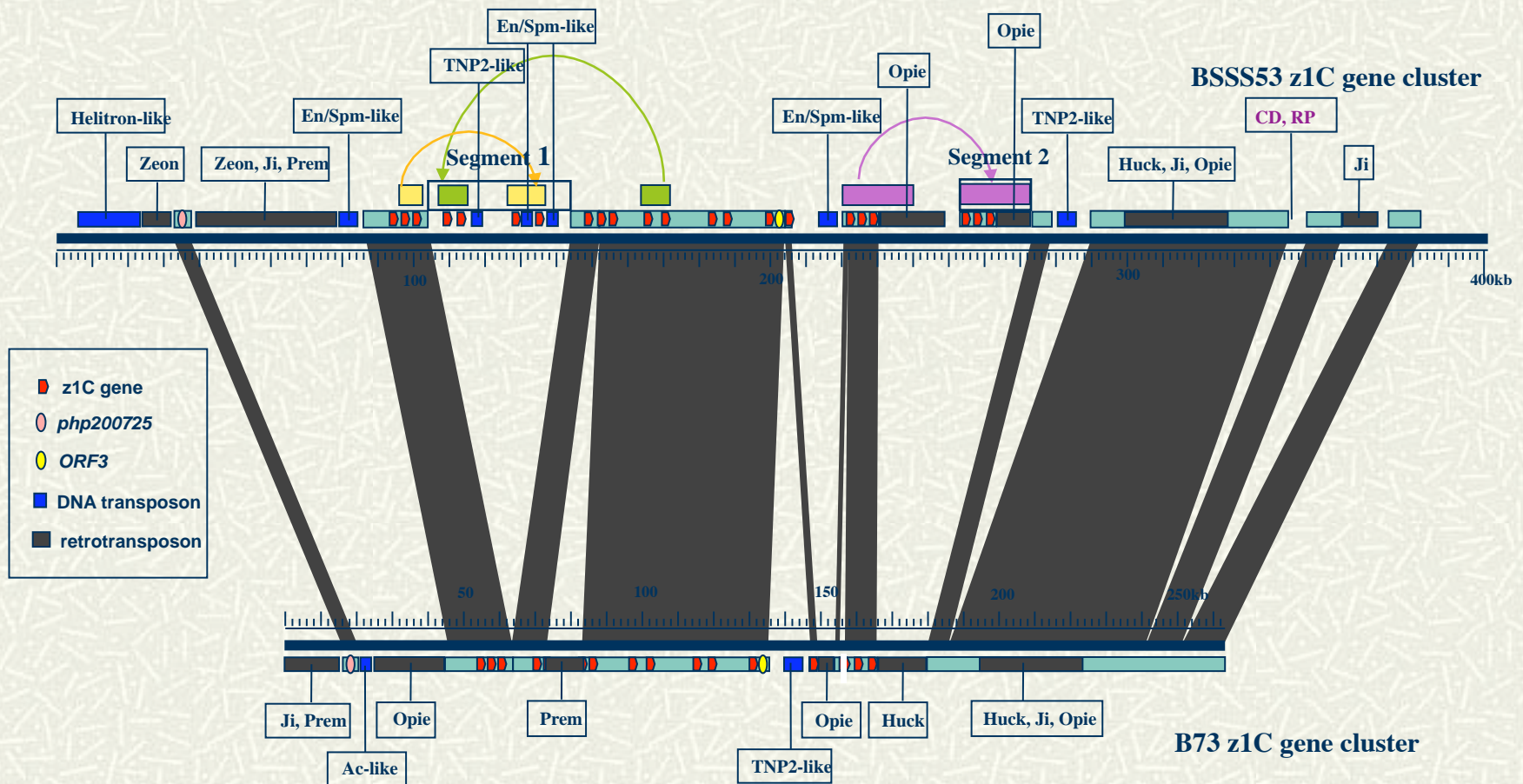
Divergence of transcriptional regulation in paralogous gene copies



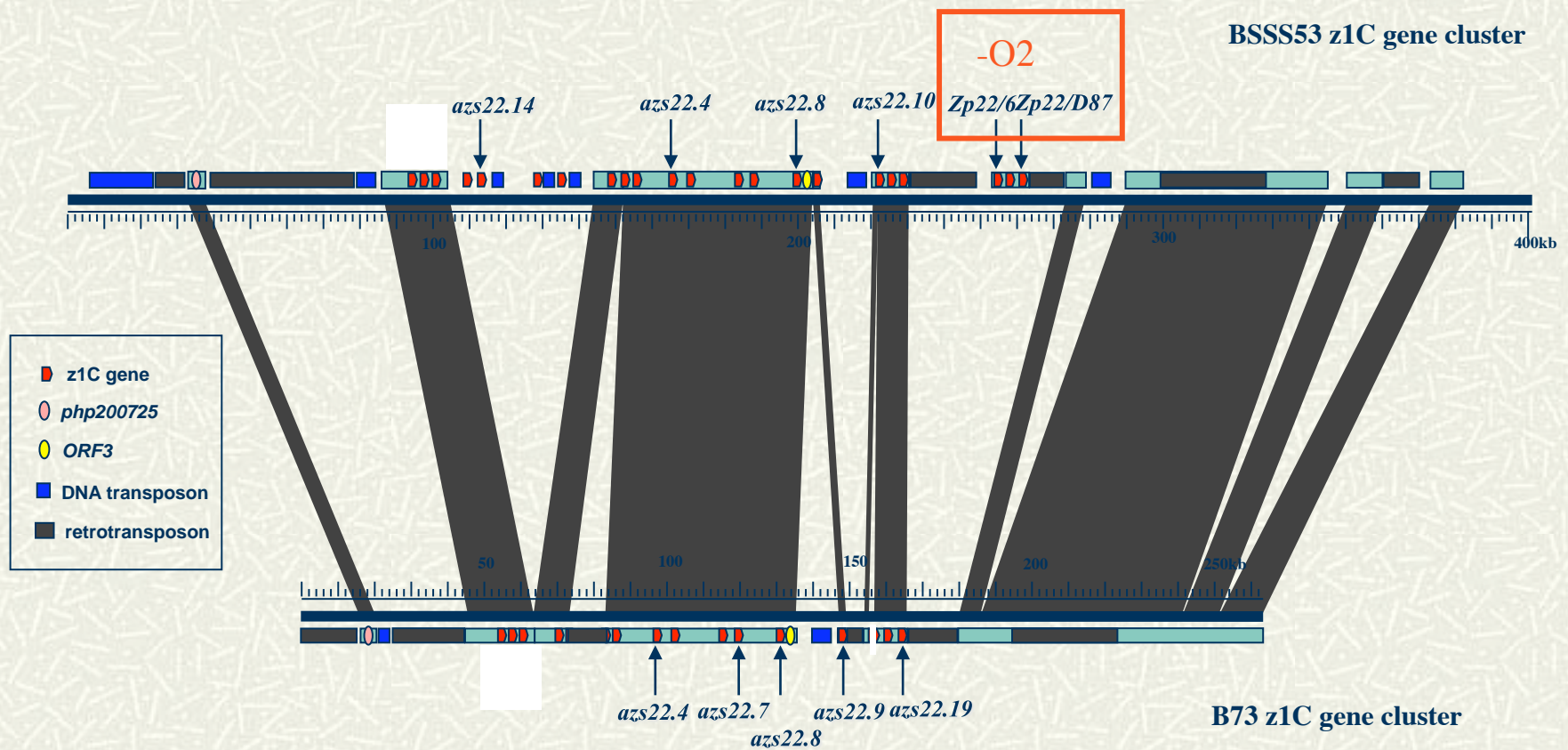
RNAi of alpha prolamin genes



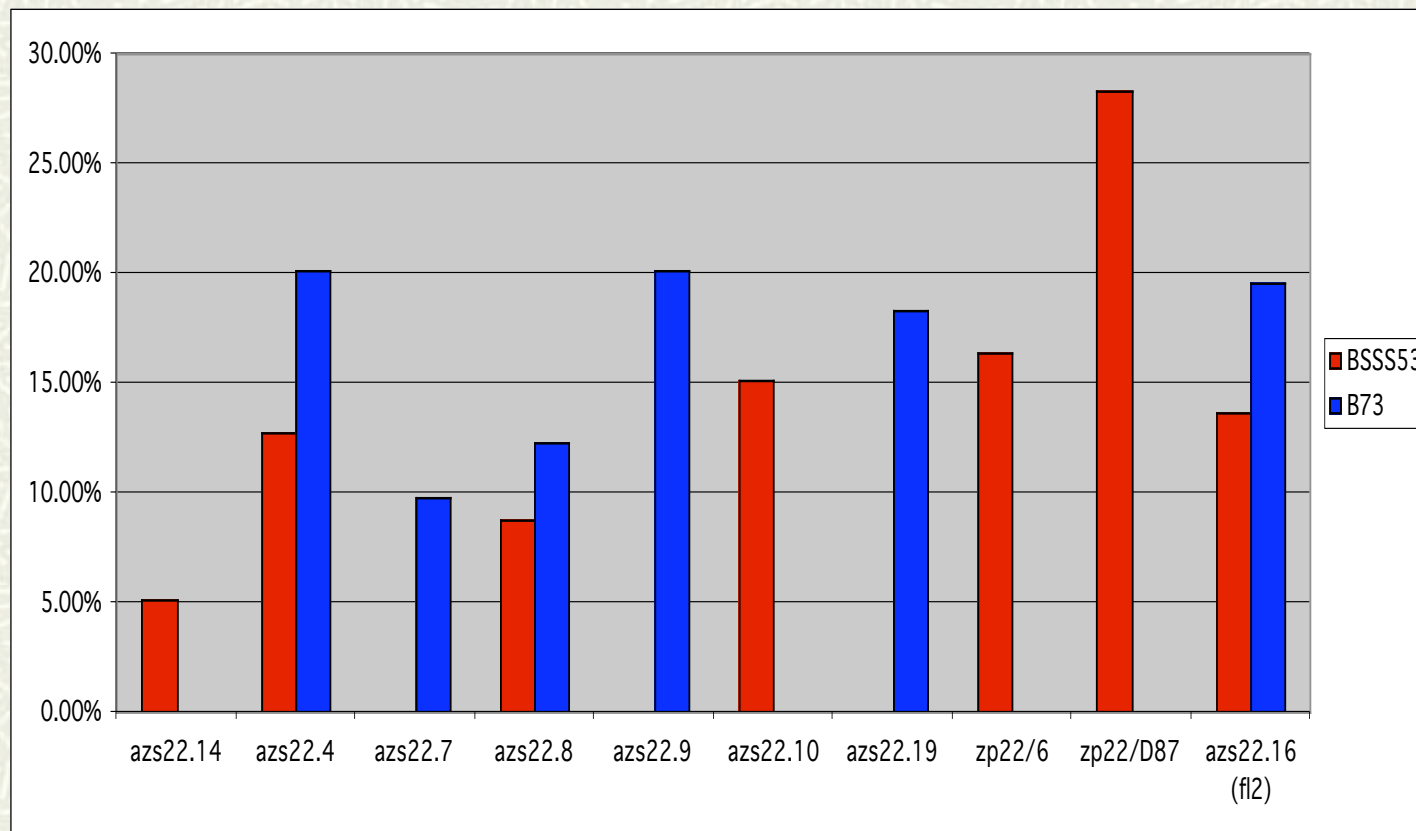
The z1C1 region on maize chromosome 4 of B73 and BSSS53 are non-collinear and differ in zein and non-zein genes



Two haplotypes with allelic and non-allelic gene expression



Relative levels of mRNA in inbred parents



Relative levels of mRNA in hybrids

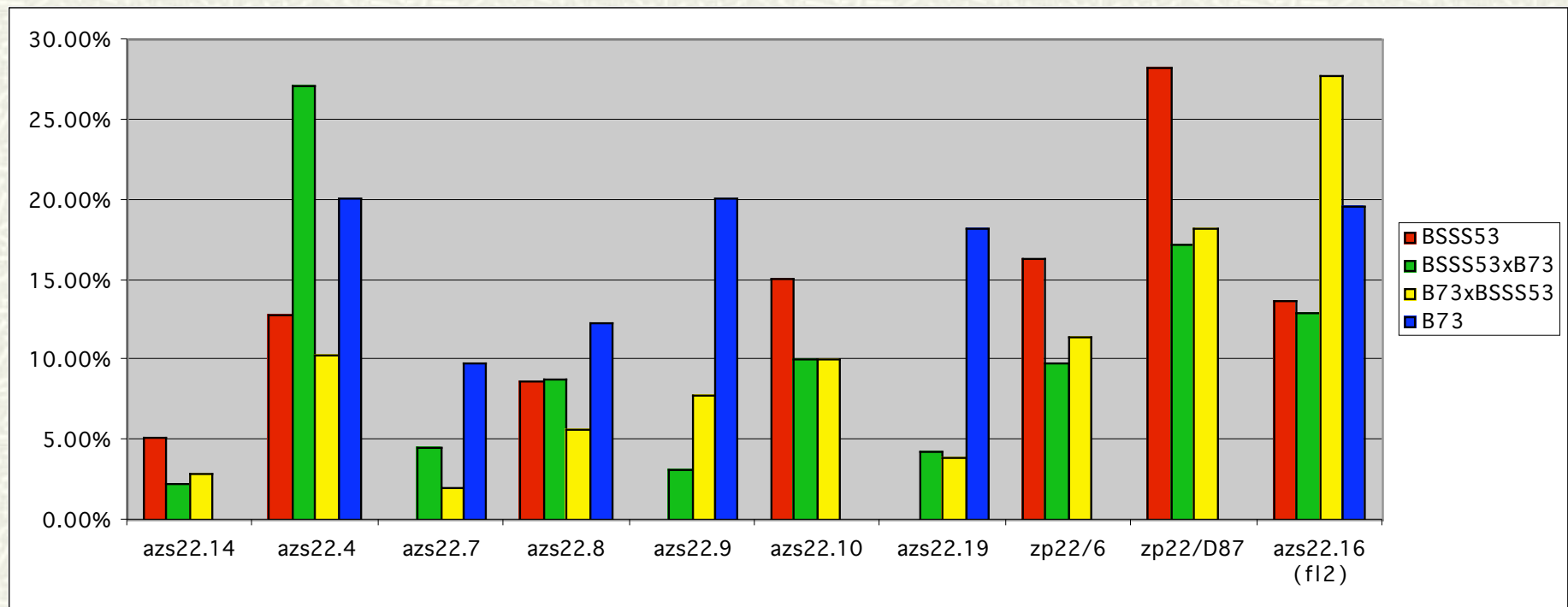
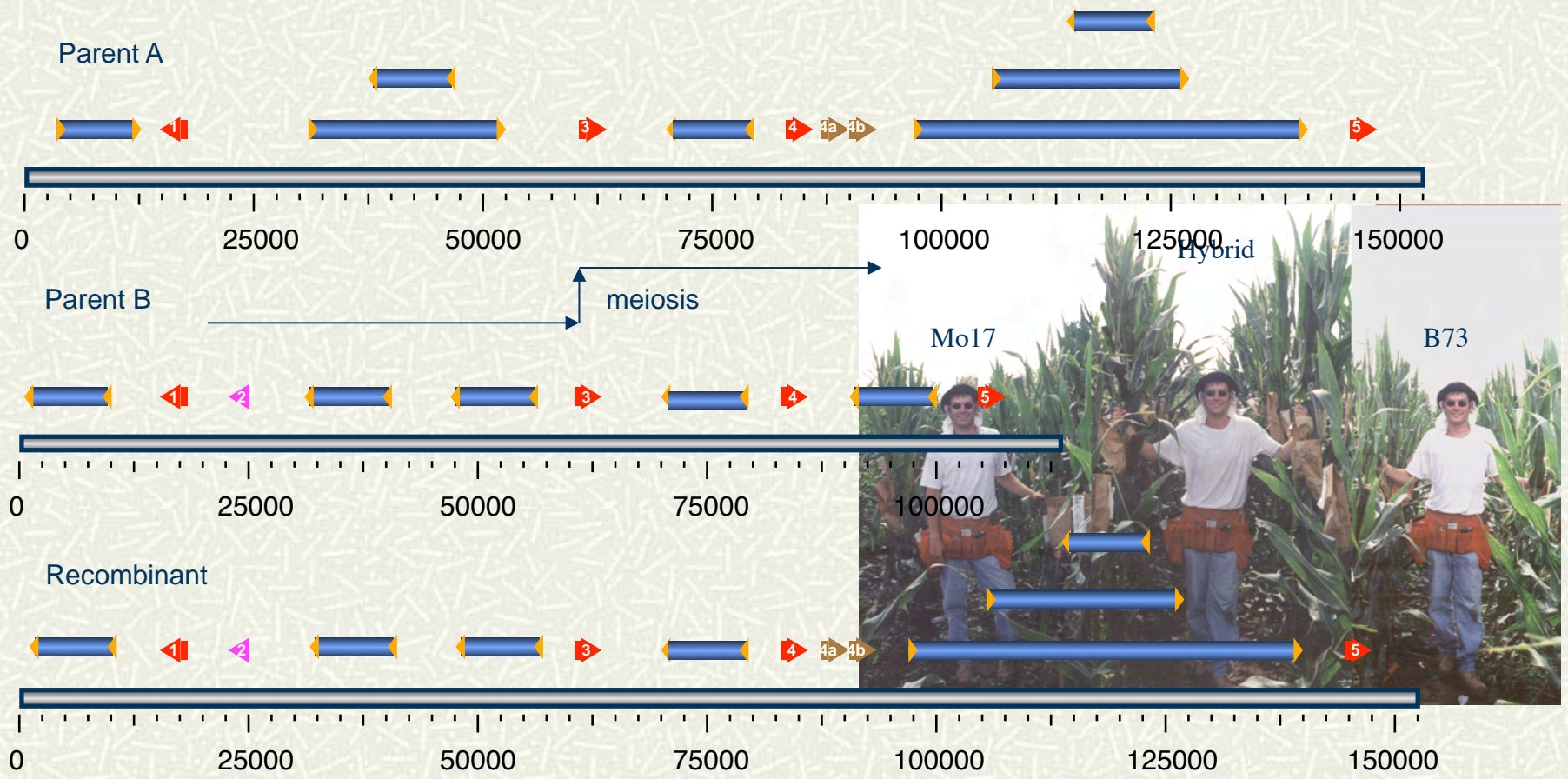


Illustration of haplotype shuffling through meiosis





Highlights

- Sequencing entire genomes has provided us with the entire set of genes of many organisms.
- Plant species appear to have very large gene sets, but vary in gene number over a large range.
- Genes are duplicated by polyploidy (orthology) or amplification (paralogy); copies of genes are also deleted and mobile.
- Gene amplification (e.g. α zein genes) can contribute to the penetrance of certain phenotypes (e.g. opaque-2); RNAi produces dominant opaque phenotype.
- Non-allelic gene copies, if expressed, provide unique properties to inbreds and hybrids because recombination occurs via allelic copies only.
- Recognizing these new features in plant genomes has important implications for genetic maps, penetrance of phenotypes, quantitative traits, and heterosis.

Acknowledgements

WEB site: <http://pgir.rutgers.edu/>

Sorghum: Remy Bruggmann

Maize origin: Zuzana Swigonova

Maize map and BACs: Arvind Bharti Galina Fuks

Segmental duplication: Jian-Hong Xu

Prolamin genes: Mihai Miclaus, Jian-Hong Xu

MIPS: Klaus Mayer, Heidrun

Gundlach, Georg Haberer

Maize map (Wing, Soderlund, Coe, Paterson)

Fusheng Wei, Will Nelson, Hye-Ran Kim, Yeisoo Yu

Sorghum (JGI /UG)

Rokshar, Paterson, Mayer

<http://www.phytozome.net/cgi-bin/gbrowse/sorghum/>

Nature (2009), 457(7229):551-6.

Support:

DOE, NSF

**Selman A. Waksman Chair
in Molecular Genetics**

