UNIVERSITÀ
DEGLI STUDI
DI UDINE

Estimating the rates and modes of creation of new genetic variation in plants using NGS technologies

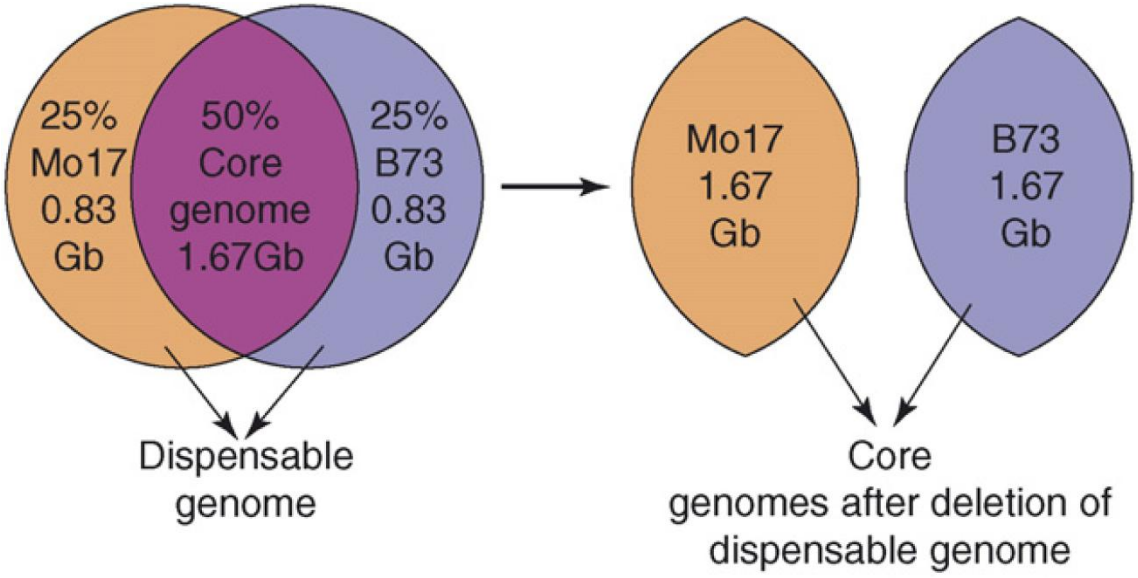
14/06/2016

Supervisor: Prof. Michele Morgante
Co-supervisor: Fabio Marroni

PhD Student: Ettore Zapparoli

The maize pan-genome

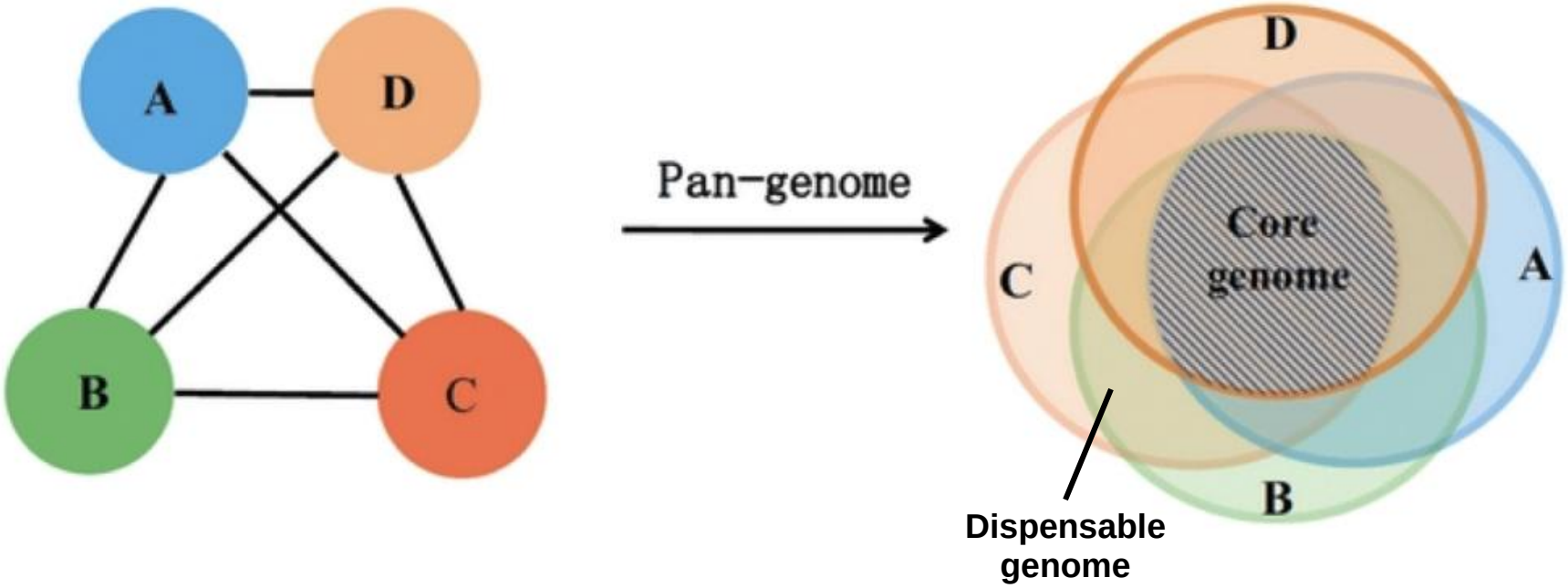
- A single individual genome is not representative of an entire species
- Pan-genome: the set of all genomic features among individuals within a species



Morgante et al., Curr Opin Plant Biol. 2007

Aim of the project

Characterizing maize dispensable genome



Modified from Novogene Corp., 2015

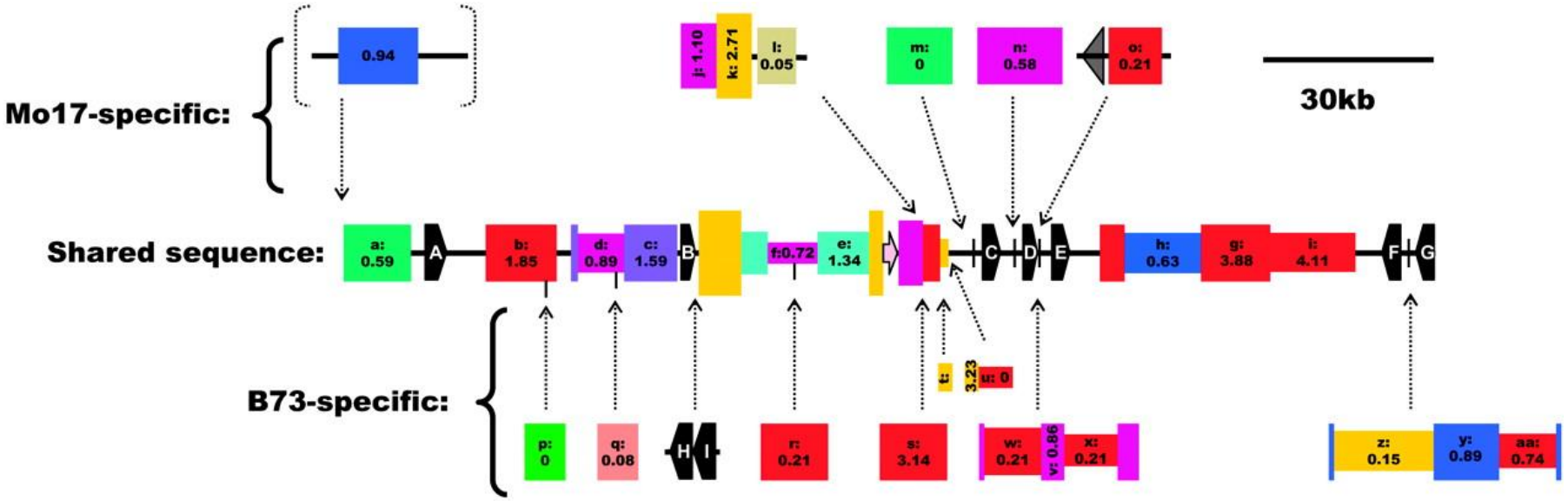
Structural variants in plants

- **Deletions** and **insertions** of TEs

- **CNV**: Copy-Number Variants

- **PAV**: Presence-Absence Variants

Insertions and deletions in maize



Genes: A) – I): geneA9008 – geneI9008

Repetitive elements:

1) Transposon:

DOPA-like

2) Retrotransposons:

- *ji*
- *opie*
- *reiver*
- *cinful*
- *flip*
- *prem*
- *huck*
- *tekay*
- *zeon*
- *shadowspawn*
- *non-LTR*

Brunner et al. Plant Cell 2005

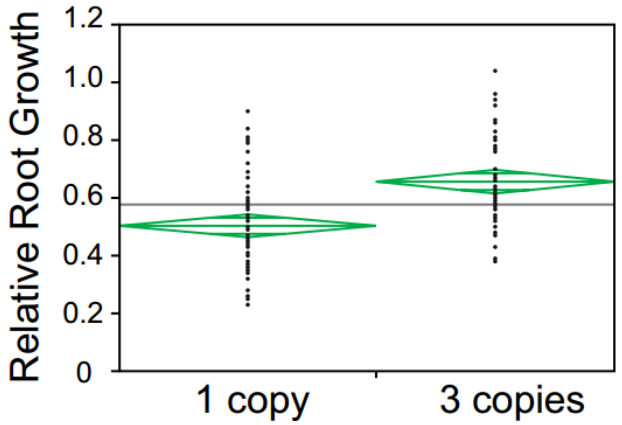
CNV in maize

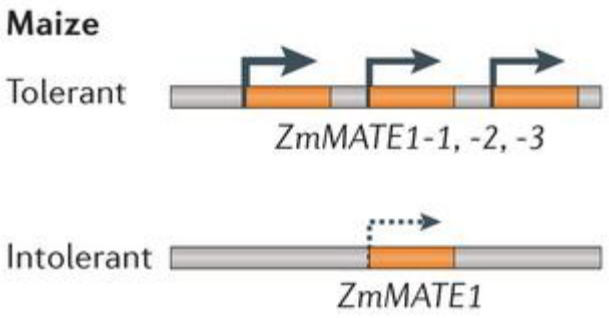
PNAS

Aluminum tolerance in maize is associated with higher *MATE1* gene copy number

Lyza G. Maron[a,1], Claudia T. Guimarães[b], Matias Kirst[c,d], Patrice S. Albert[e], James A. Birchler[e], Peter J. Bradbury[a,f], Edward S. Buckler[a,f], Alison E. Coluccio[a], Tatiana V. Danilova[e,2], David Kudrna[g], Jurandir V. Magalhaes[b], Miguel A. Piñeros[a], Michael C. Schatz[h], Rod A. Wing[g], and Leon V. Kochian[a,1]

[a]Robert W. Holley Center for Agriculture and Health, US Department of Agriculture–Agricultural Research Service, and [f]Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853; [b]Embrapa Maize and Sorghum, 35701-970, Sete Lagoas, Minas Gerais, Brazil; [c]School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611; [d]University of Florida Genetics Institute, University of Florida, Gainesville, FL 32610; [e]Division of Biological Sciences, University of Missouri, Columbia, MO 65211; [g]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85705; and [h]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

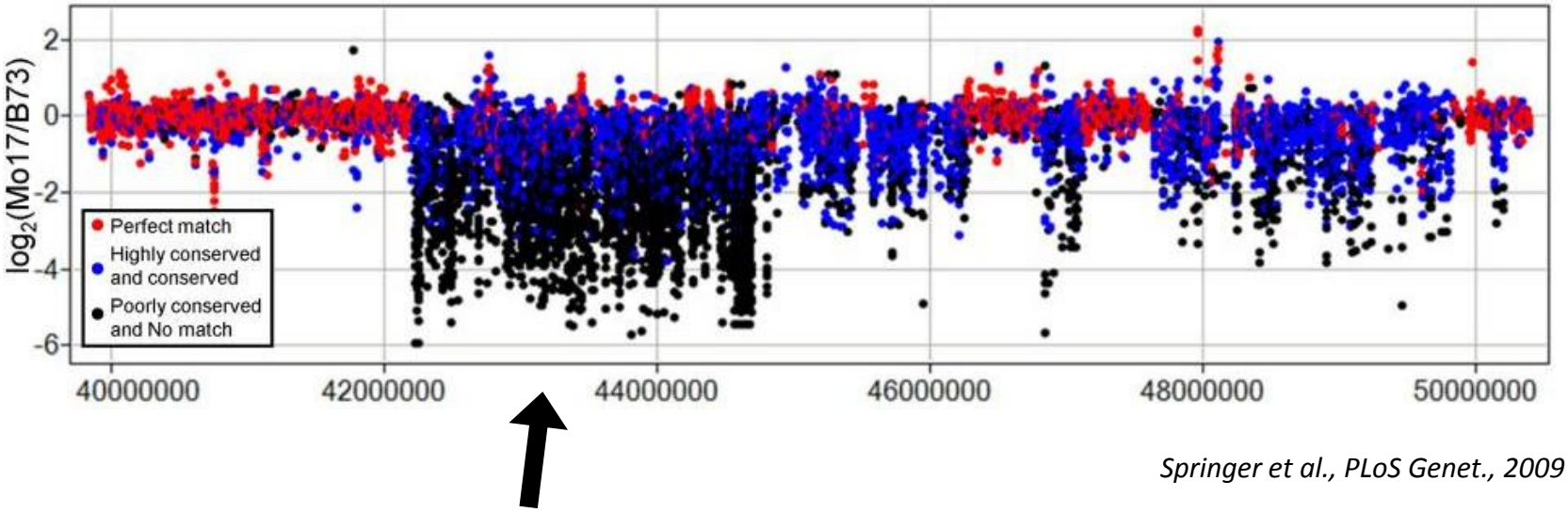




Mickelbart et al, Nature Rev Gen, 2015

PAV in maize

► ***Short arm Chr 6***

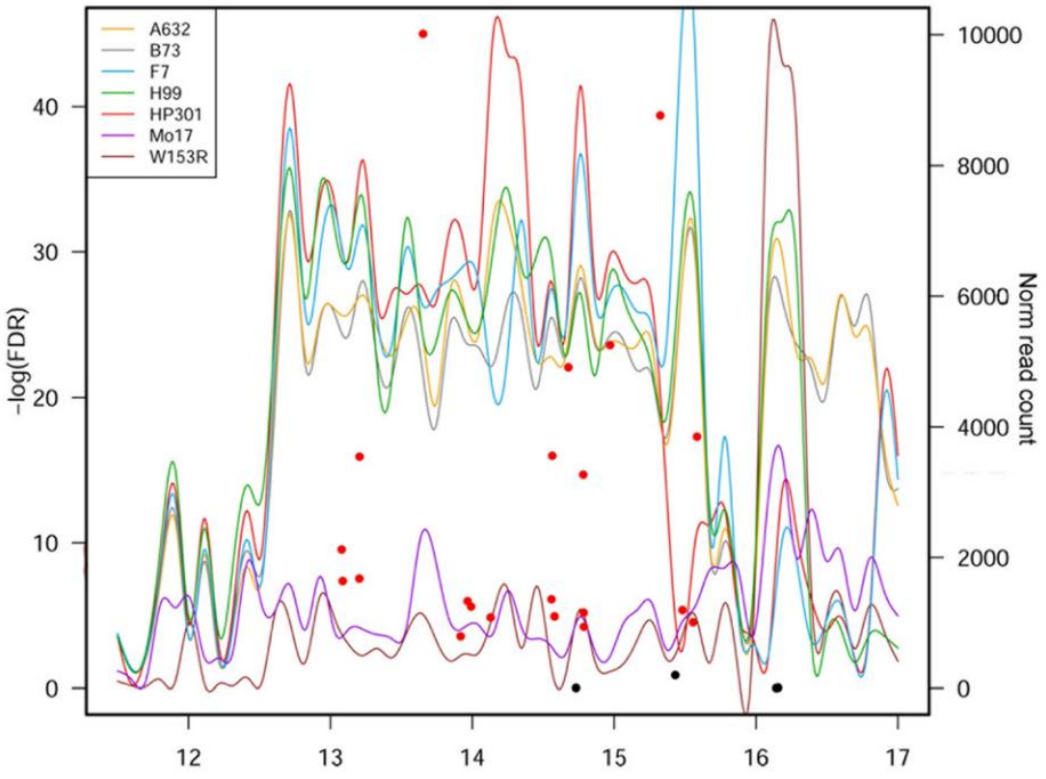


Springer et al., PLoS Genet., 2009

► ***2 Mb present in B73 and absent in Mo17***

PAV in maize

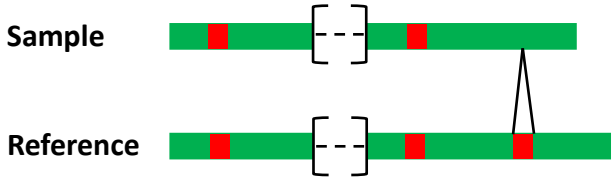
► ***Short arm Chr 6***



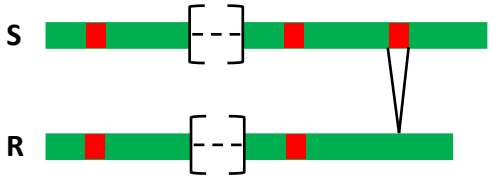
Dell'Acqua et al., Genome Biology, 2015

Structural variants discovery

1. Discovering **deletions**



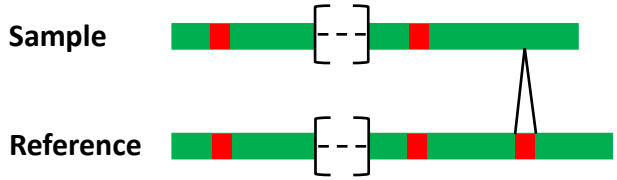
2. Discovering **insertions**



3. Discovering **CNV**s

Structural variants discovery

1. Discovering **deletions** (< 50 kbp)

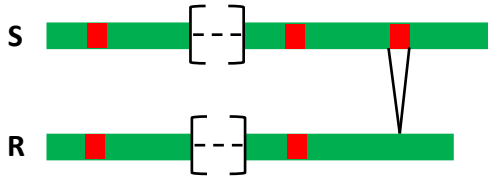


➔ Applying available methods

(paired-end mapping)

Structural variants discovery

2. Discovering **insertions** (< 50 kbp)



- Applying available methods

  (paired-end mapping)

- Developing new methods

  (split-reads mapping)

Structural variants discovery

3. Discovering **CNV**s (> 50 Kbp)

➔ Applying available methods

(depth of coverage approach)

Validation of detected deletions

- Integrated pipeline *(Magris, 2016)* calling:
  - → Delly *(Rausch et al., Bioinformatics 2012)*
  - → GASV *(Sindi et al., Bioinformatics 2009)*
- Validation in maize:
  - → BAC sequences (Mo17; >1 Mb analyzed, 33 events)
  - → genome assembly drafts (H99, HP301; 53 715 events)

Precision	Sensitivity	F1 score
0.92	**0.66**	**0.77**

Validation of detected insertions

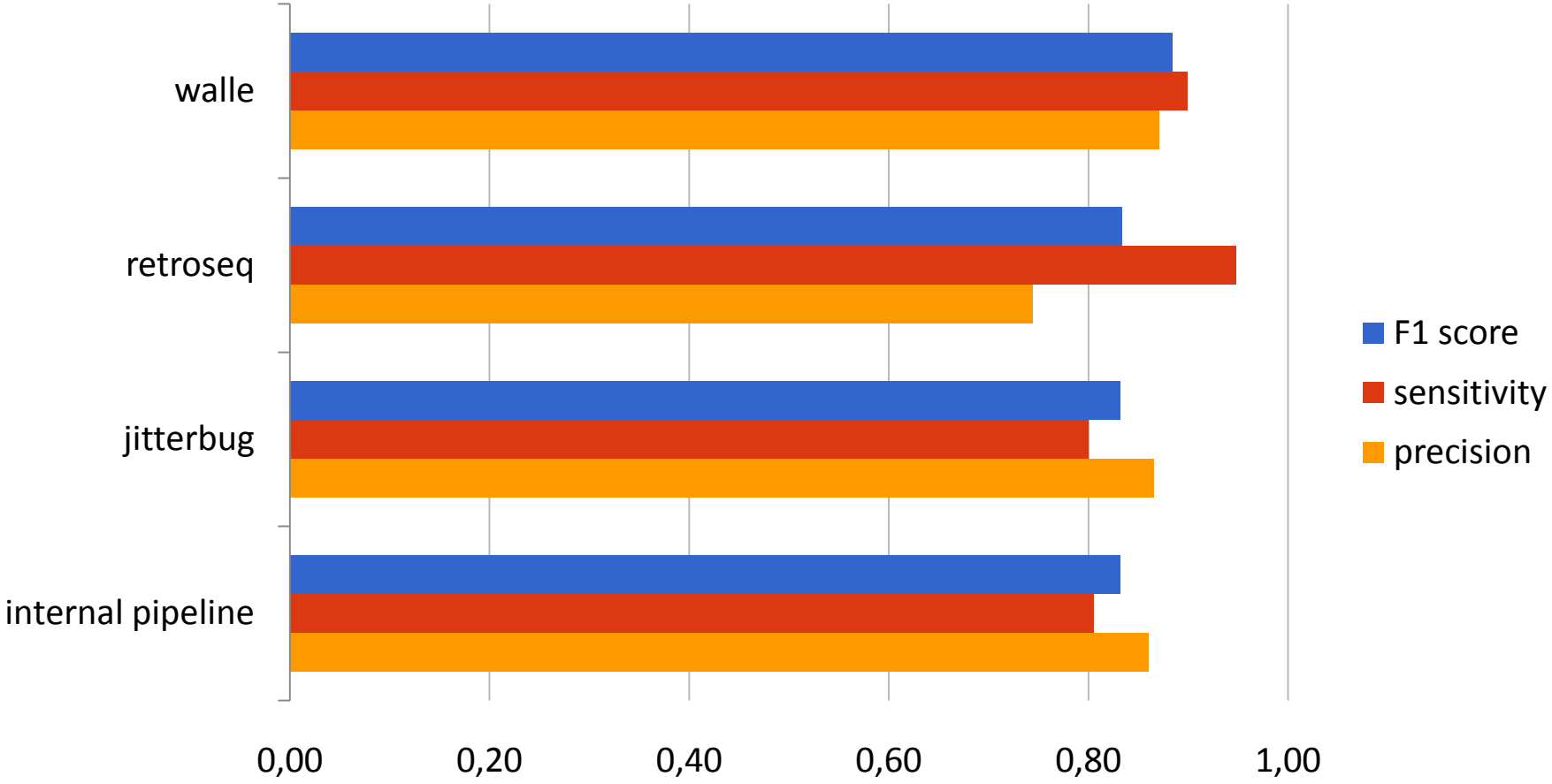
- TE DB-dependent methods
  - → IGA internal pipeline *(Pinosio, 2012)* - PEM, De Novo assembly
  - → Jitterbug *(Hénaff et al., BMC Genomics 2015)*
  - → RetroSeq *(Keane et al., Bioinformatics 2012)*
- Validation in maize:
  - → BAC sequences (Mo17; >1 Mb analyzed, 21 events)
  - → genome assembly drafts (in progress)
- Preliminary analysis on BACs reveals an overall good power (F1>0.8)

Developing Walle

- Developing a TE DB independent method: Walle
  - → Split-reads mapping approach
  - → Able to find insertions of not annotated elements

- Validation in vitis (now running in maize):
  - → Simulation of 1000 insertions
  - → Outperformed direct competitors

Validation of Walle

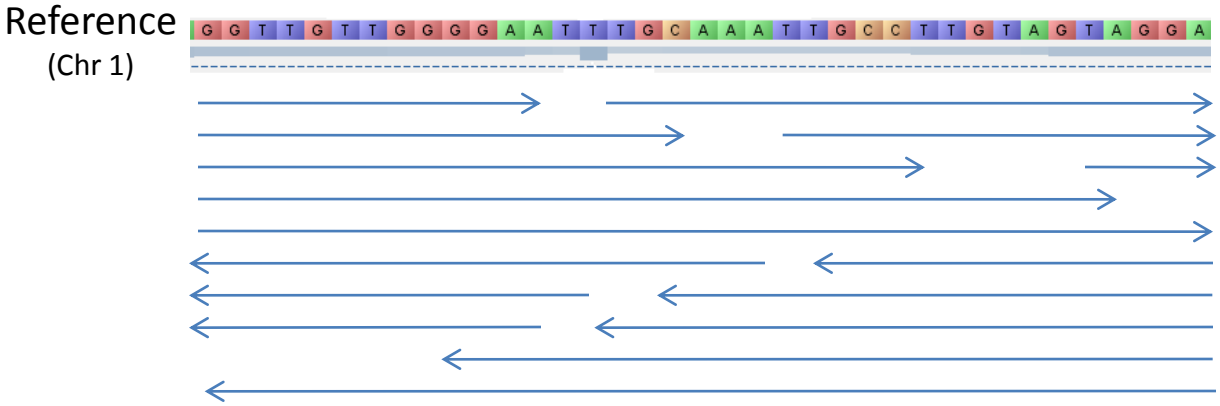


Walle methods

Split-reads mapping approach

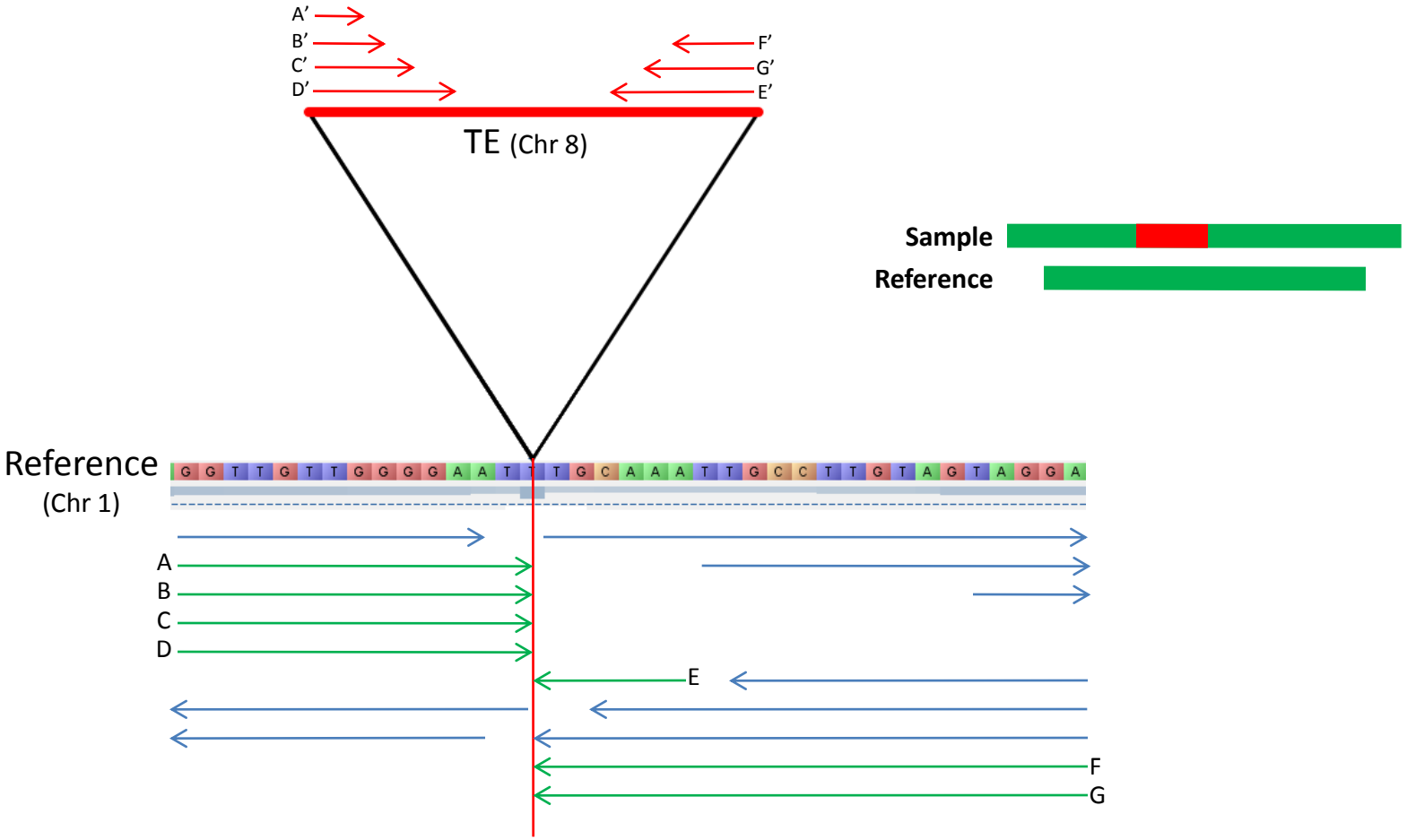
Normal alignment

(no insertion events)

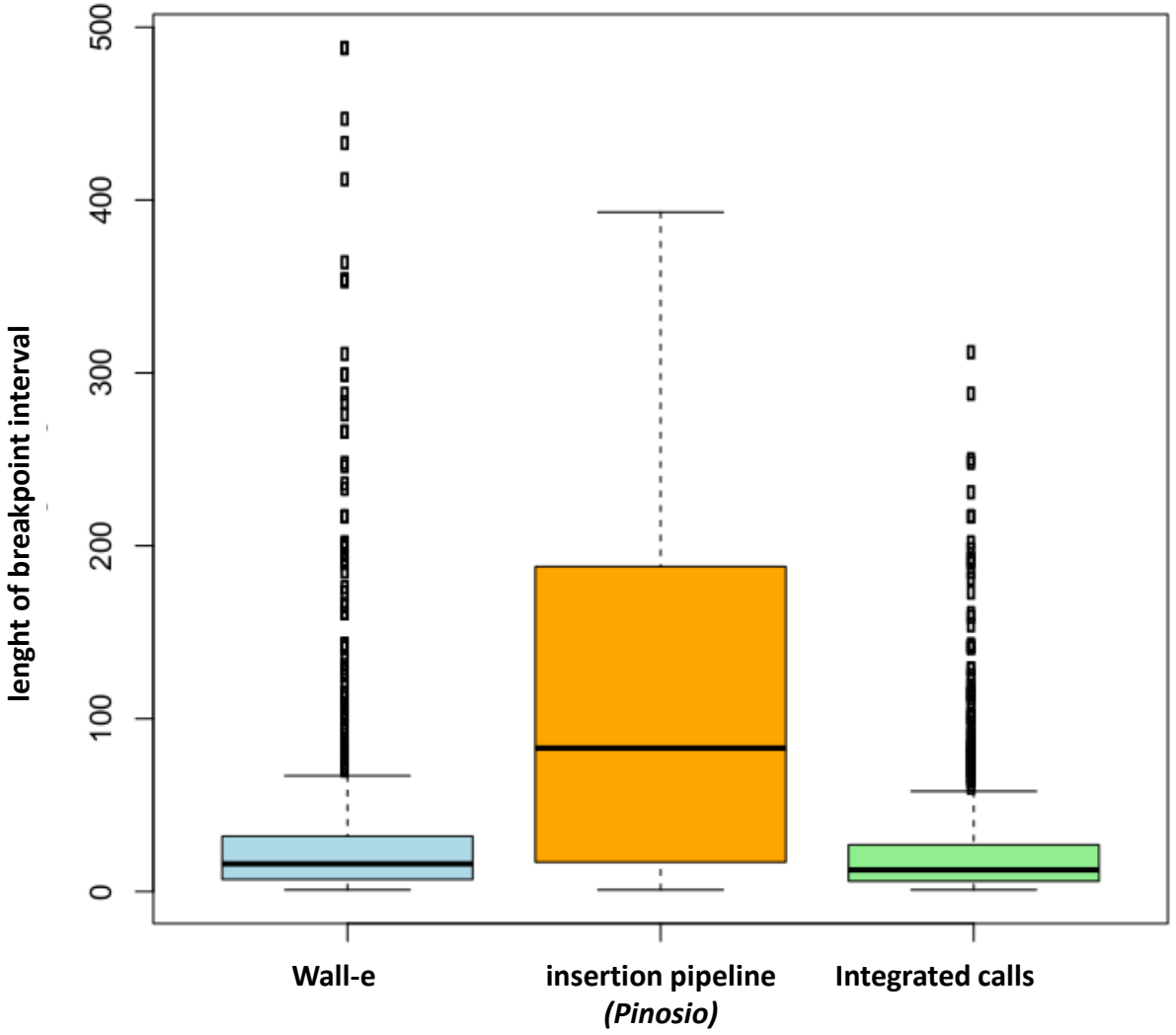


Walle methods

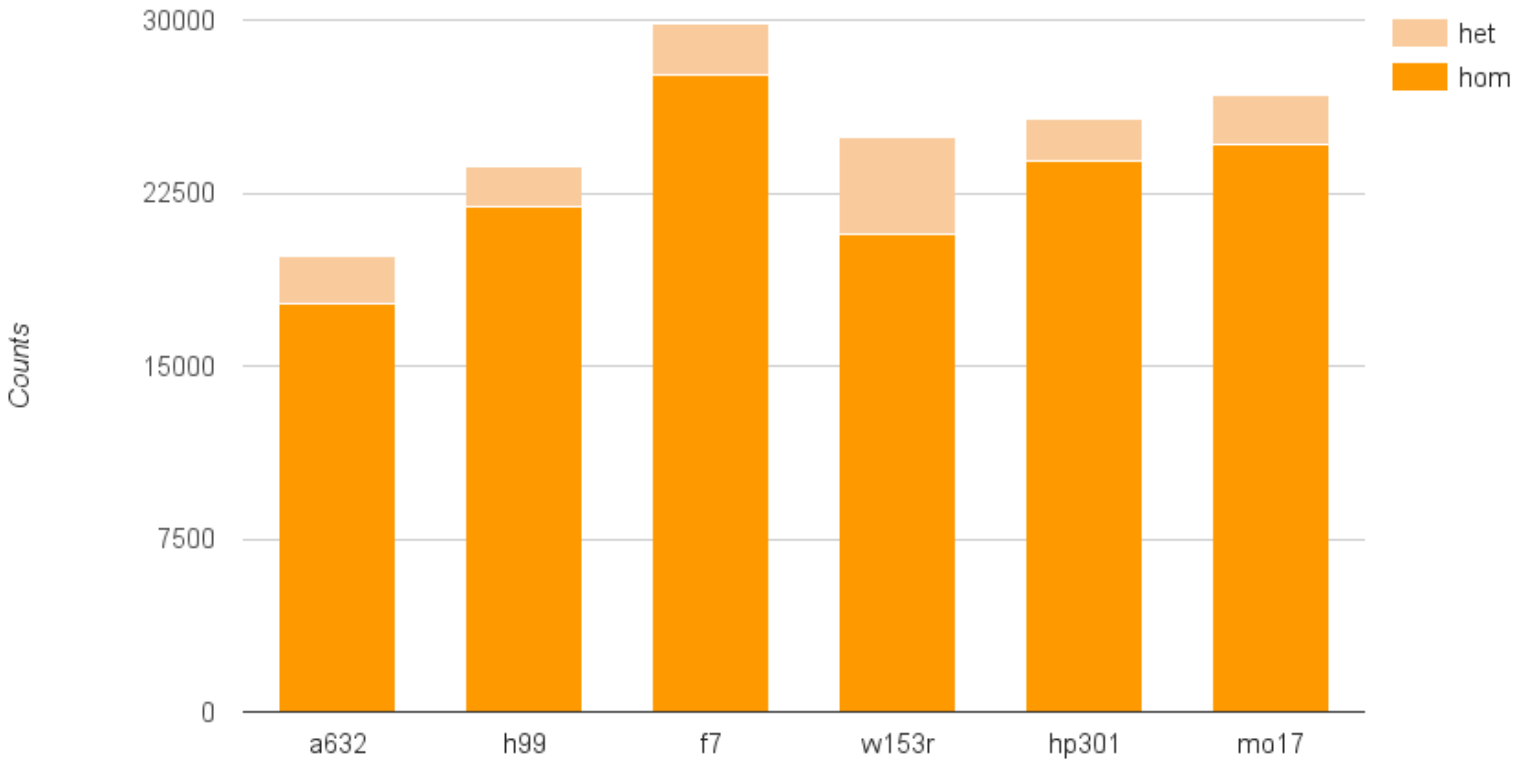
Split-reads mapping approach



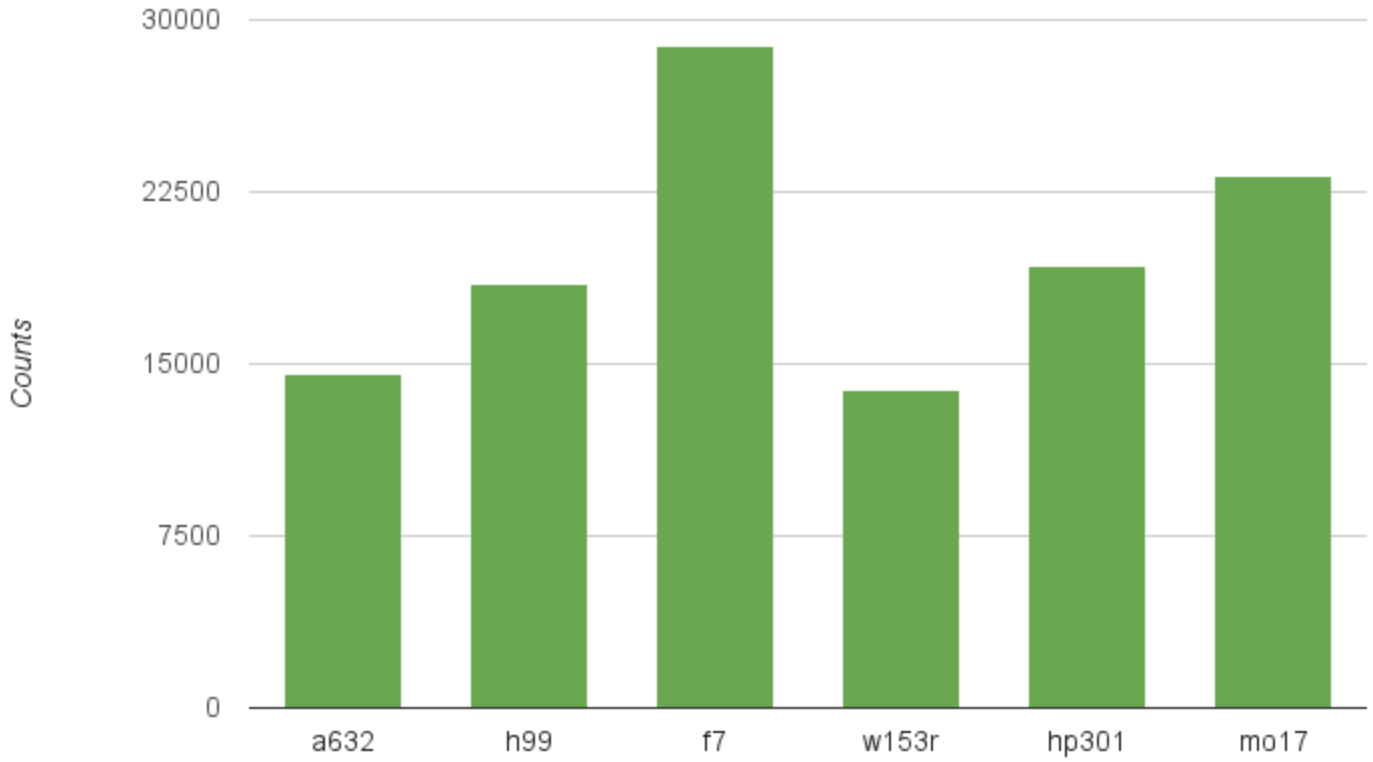
Refinement of IGA internal tool calls



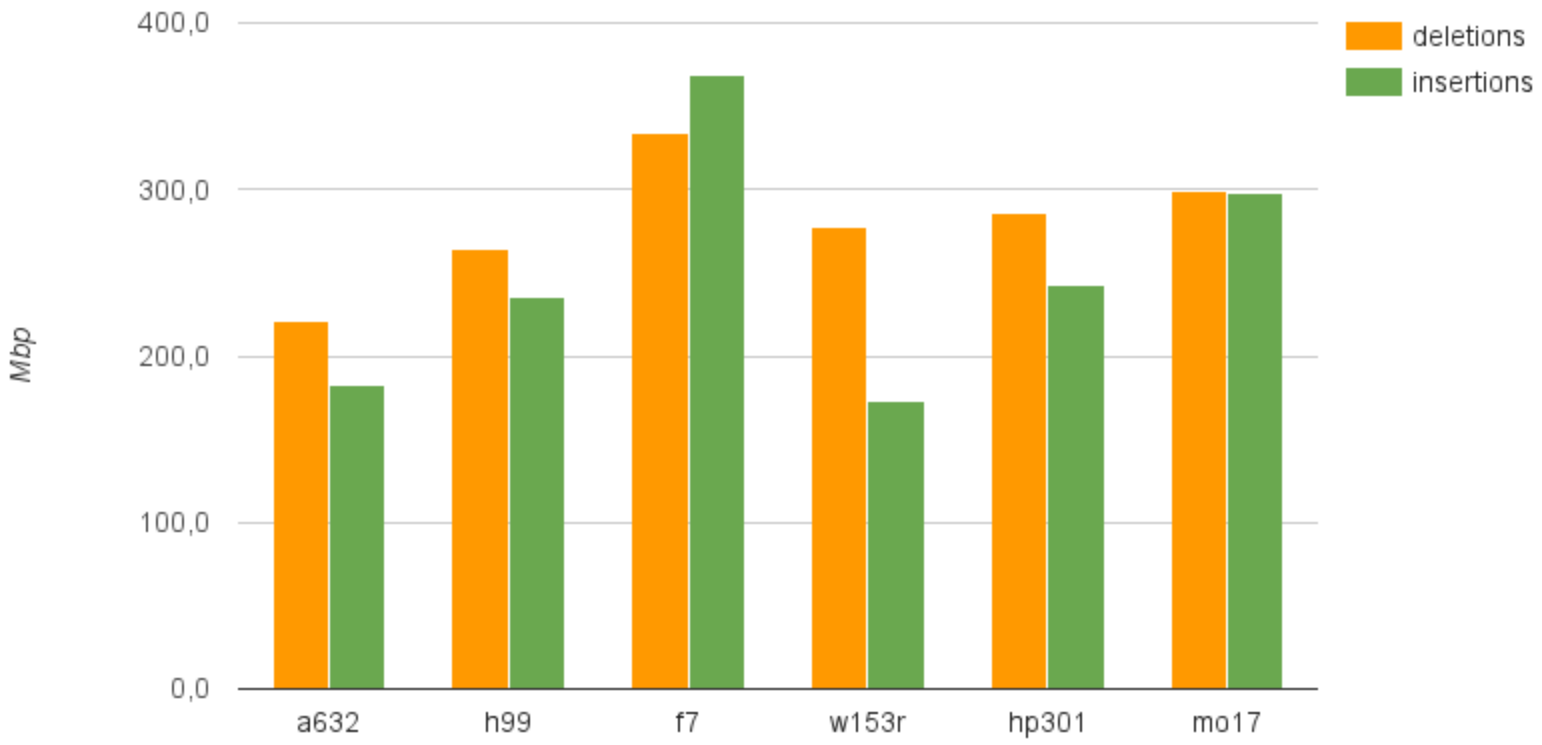
Deletions in maize



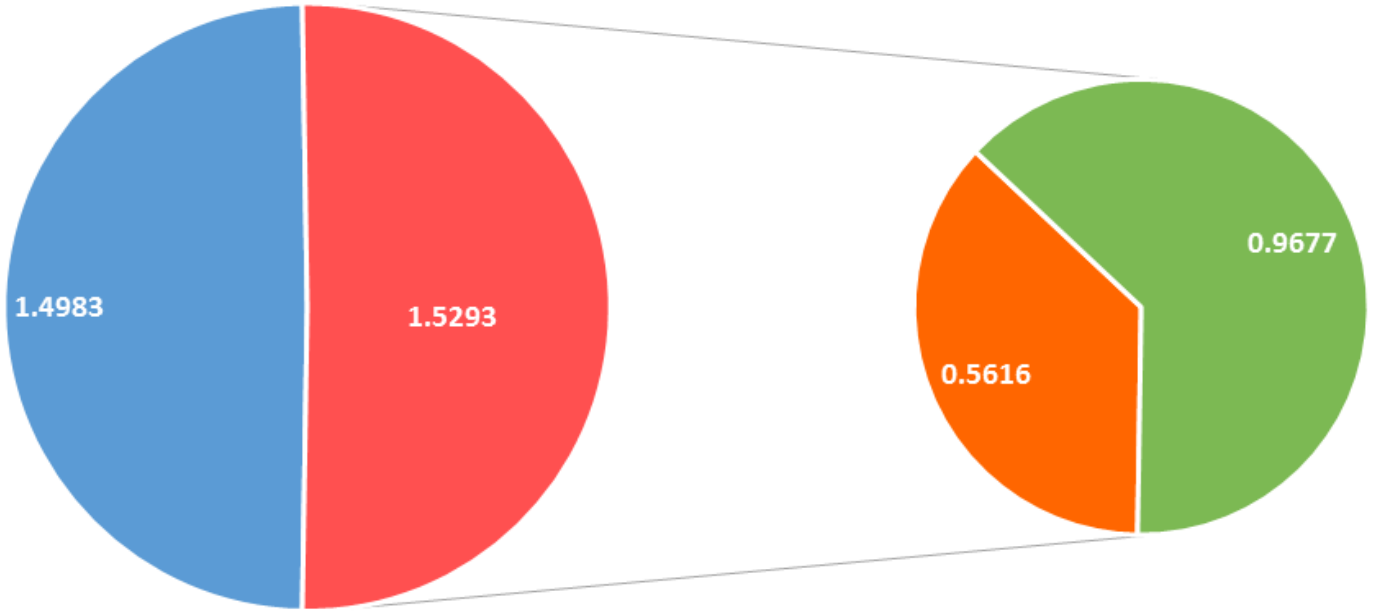
Insertions in maize



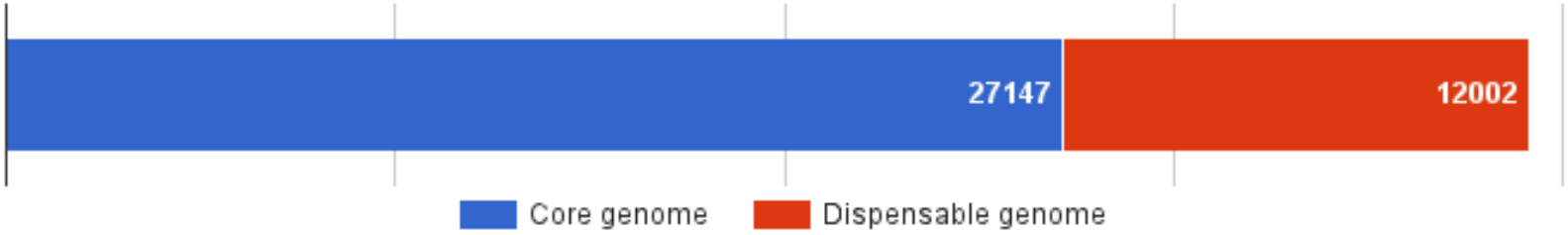
Size of deletions and insertions



Maize pan-genome



Shared genes analysis



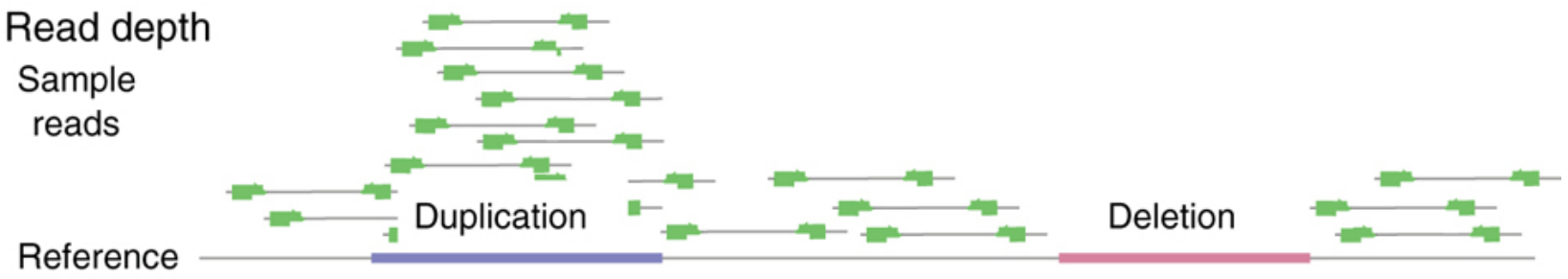
- Analysis performed on 7 varieties (included B73)

- **PanViz** *D3 based visualization for comparative genomics*

  (https://github.com/thomasp85/PanViz - *Pedersen TL, 2016*)

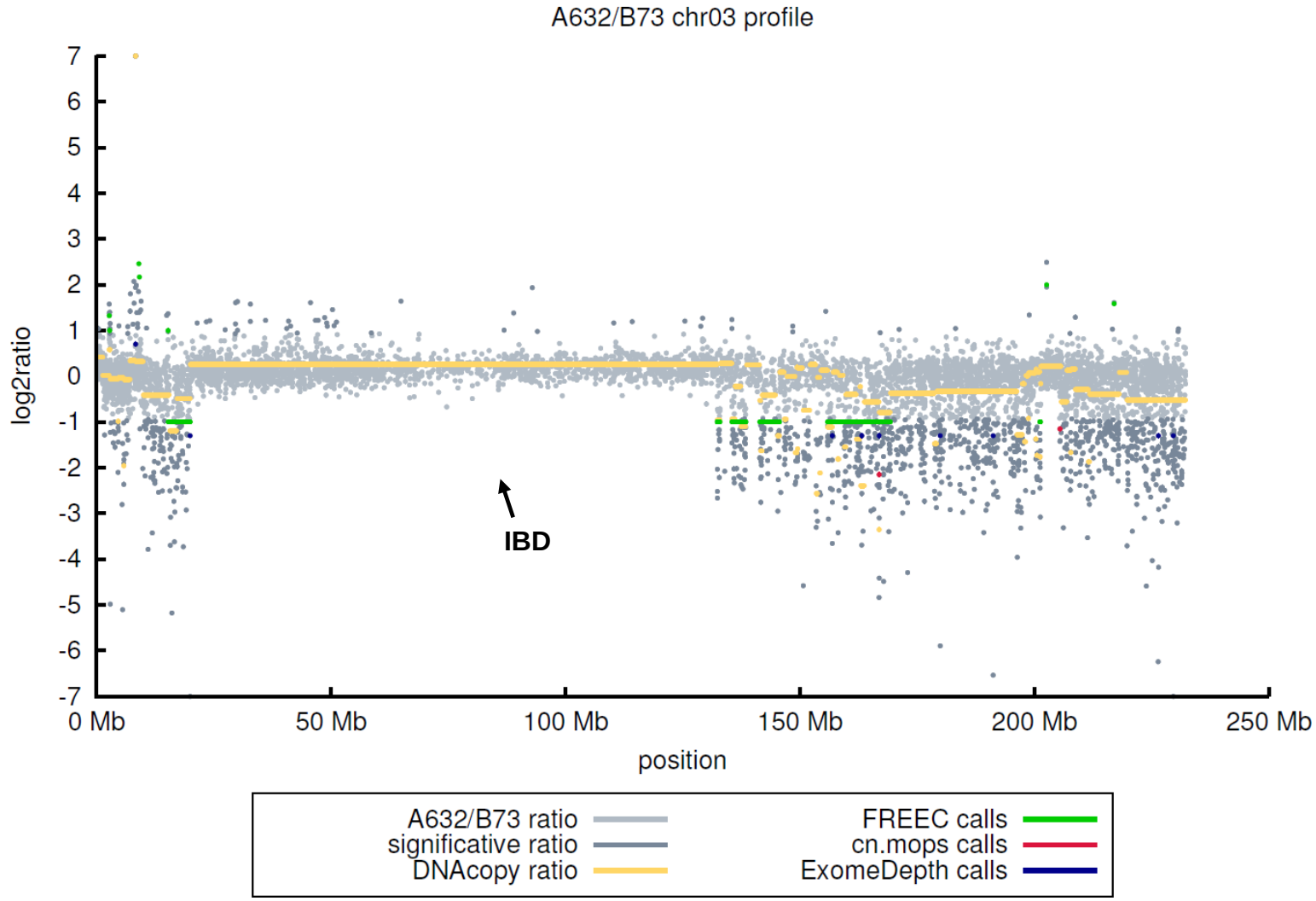
Discovering CNV

- Computing the ratio between the read counts of a sample genome and a reference genome in a region

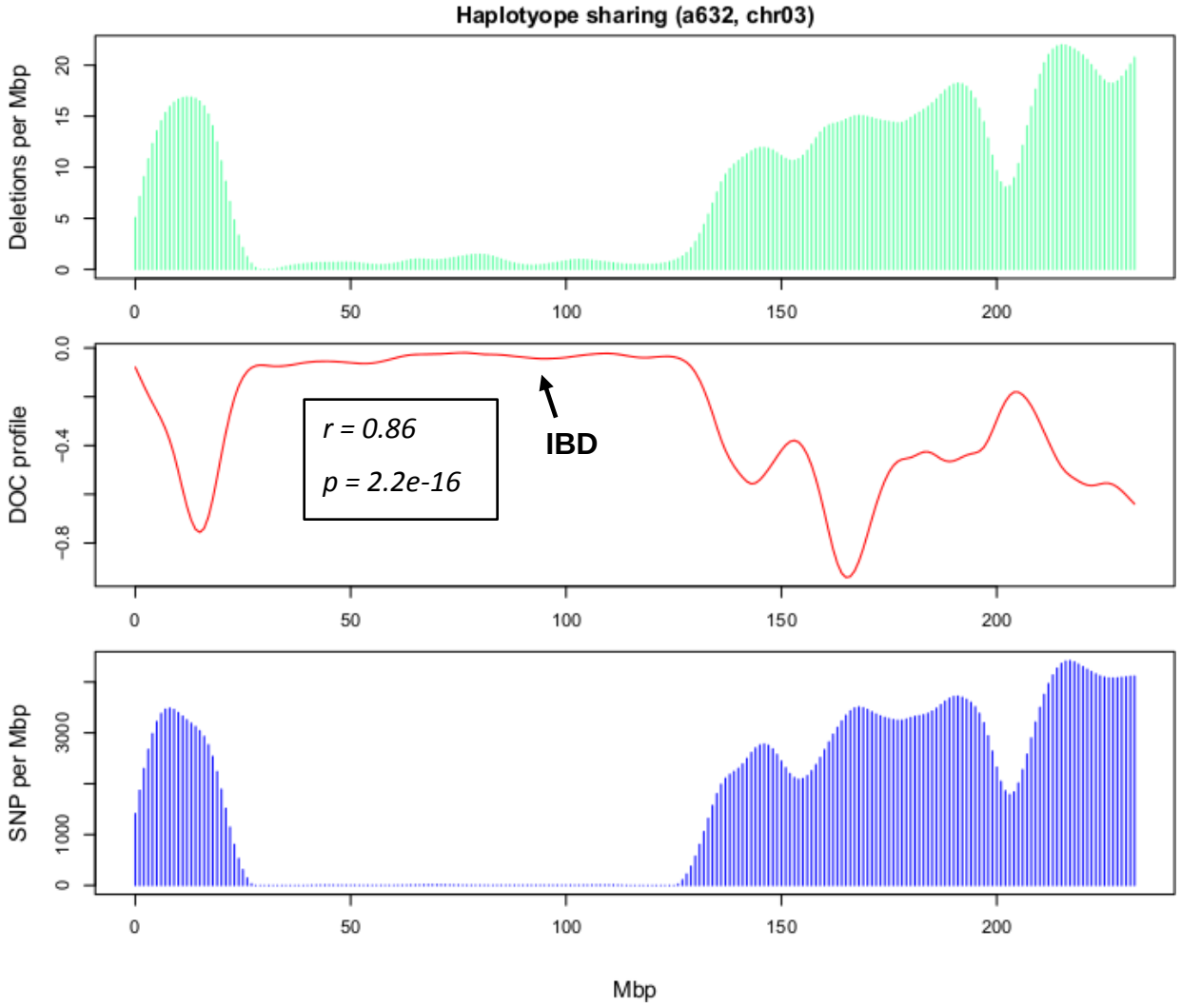


(Baker 2012)

DOC profiles



Correlating deletions and DOC



Future perspectives

- Insertions validation and genotyping
  - » Finalize Walle and validate it in maize
- Obtaining a list of CNVs in maize
- Complete the characterization of maize dispensable genome
  - » Obtaining a list of involved genes
  - » Identifying TE classes more often involved in SV
  - » Dating of amplification events of LTR sequences

Acknowledgments

Michele Morgante

Fabio Marroni

Gabriele Magris

Michele Vidotto

Davide Scaglione

Mirko Celii

Sara Pinosio

Vittorio Zamboni

Cristian Del Fabbro

Alice Fornasiero

Aldo Tocci

Mara Miculan

and all the IGA crew.

Supported by the ERC project NOVABREED - Novel variation in plant breeding and the plant pan-genomes (Grant agreement no.: 294780).

Partial Bibliography

Marroni et al.

Structural variation and genome complexity: is dispensable really dispensable?

Curr Opin Plant Biol, 2014

Morgante et al.

Transposable elements and the plant pan-genomes

Curr Opin Plant Biol, 2007

Brunner et al.

Evolution of DNA sequence nonhomologies among maize inbreds

Plant Cell, 2005