

# Novabreed

*Update on the project status*

Fabio Marroni



European Research Council

Established by the European Commission

# Work packages

- **Workpackage 1:** Reconstructing the pan-genome in two plant species, maize and grapevine.
- **Workpackage 2:** Identifying the different mechanisms that generate and maintain the dispensable portion in the same two species.
- **Workpackage 3:** Identifying the phenotypic effects of the dispensable genome.
- **Workpackage 4:** Estimating the rates and modes of creation of new genetic variation due to dispensable genome components and its contribution to breeding progress.
- **Workpackage 5:** Extend our findings to other plant species for which the genome sequence in the meantime may have become available.

# WP1

- Pan-genome reconstruction
  - De-novo assembly (Michele Vidotto)
  - Methods for detection of SVs (Sara Pinosio, Gabriele Magris, Ettore Zapparoli)

# De-novo assembly strategy

- Mainly based on short reads
  - a) ~30X Overlapping fragment libraries
  - b) ~20X Non-overlapping fragment libraries
  - c) ~20X **3kb** Jumping libraries + ~10X **10kb** (long) Jumping libraries
- PacBio long reads (ongoing)

# De-novo results: grape

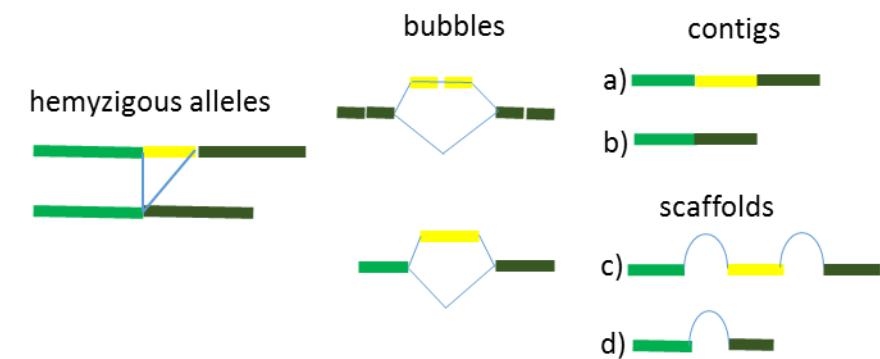
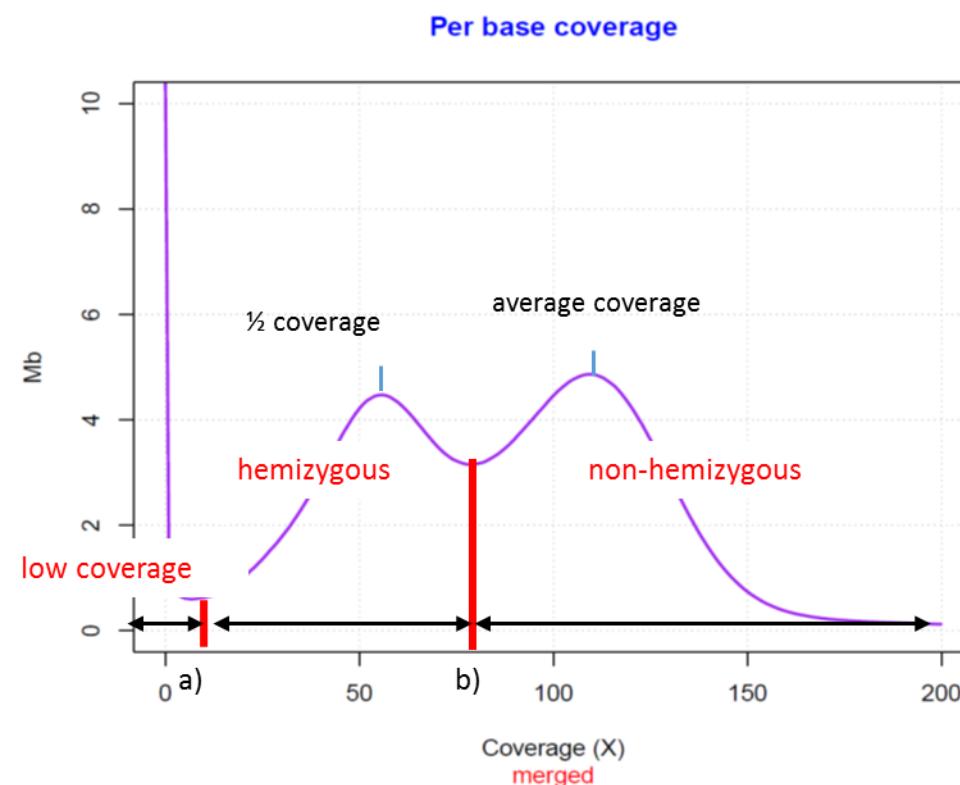
|                            | traminer    |             | gouais blanc |             | cabernet franc |             | pn40024     |             |
|----------------------------|-------------|-------------|--------------|-------------|----------------|-------------|-------------|-------------|
| Estimated Genome size (bp) | 394,441,994 |             | 420,689,048  |             | 483,072,260    |             | 444,716,040 |             |
| Estimated CN=1 (bp)        | 71.1%       |             | 69.6%        |             | 59.6%          |             | 65.0 %      |             |
| Estimated CN>1 (bp)        | 28.9%       |             | 30.4%        |             | 40.4%          |             | 35.0 %      |             |
| Coverage estimated (X)     | 53          |             | 58           |             | 50             |             | 64          |             |
|                            | contigs     | scaffolds   | contigs      | scaffolds   | contigs        | scaffolds   | contigs     | scaffolds   |
| Total (bp)                 | 412,472,184 | 678,662,631 | 415,882,525  | 743,730,069 | 487,795,456    | 711,707,043 | 377,475,581 | 521,491,794 |
| Average (bp)               | 3,728       | 78,223      | 3,722        | 77,764      | 4,770          | 57,005      | 6,814       | 86,684      |
| Max (bp)                   | 137,674     | 2,817,510   | 113,144      | 2,522,649   | 180,444        | 1,496,800   | 148,414     | 1,727,814   |
| Min (bp)                   | 32          | 885         | 69           | 905         | 58             | 894         | 118         | 885         |
| Sequences (#)              | 110,653     | 8,676       | 111,742      | 9,564       | 102,258        | 12,485      | 55,399      | 6,016       |
| N50 (#)                    | 12,360      | 484         | 12,724       | 524         | 9,708          | 1,178       | 7,151       | 553         |
| L50 (bp)                   | 7,841       | 413,025     | 7,541        | 408,071     | 11,145         | 178,391     | 14,075      | 264,099     |
| Gaps (#)                   |             | 67,484      |              | 66,217      |                | 55,190      |             | 40,378      |
| Median gap length (bp)     |             | 4,012       |              | 5,545       |                | 4,405       |             | 3,825       |

|                            | rkatsiteli  |             | kishmish vatkana |             | sangiovese  |             |
|----------------------------|-------------|-------------|------------------|-------------|-------------|-------------|
| Estimated Genome size (bp) | 407,302,602 |             | 454,927,661      |             | 420,525,421 |             |
| Estimated CN=1 (bp)        | 69.1 %      |             | 56.5 %           |             | 68.4 %      |             |
| Estimated CN>1 (bp)        | 30.9 %      |             | 43.5 %           |             | 31.6 %      |             |
| Coverage estimated (X)     | 52          |             | 46               |             | 64          |             |
|                            | contigs     | scaffolds   | contigs          | scaffolds   | contigs     | scaffolds   |
| Total (bp)                 | 426,363,529 | 753,744,202 | 410,415,417      | 610,410,426 | 412,308,934 | 691,235,812 |
| Average (bp)               | 3,739       | 74,710      | 3,803            | 71,426      | 4,044       | 44,121      |
| Max (bp)                   | 135,733     | 2,436,454   | 119,547          | 2,355,113   | 184,133     | 2,528,353   |
| Min (bp)                   | 46          | 889         | 27               | 895         | 76          | 881         |
| Sequences (#)              | 114,026     | 10,089      | 107,916          | 8,546       | 101,962     | 15,667      |
| N50 (#)                    | 14,103      | 612         | 12,979           | 457         | 10,243      | 627         |
| L50 (bp)                   | 7,250       | 352,572     | 7,751            | 364,852     | 8,975       | 301,183     |
| Gaps (#)                   |             | 69,862      |                  | 64,039      |             | 61,417      |
| Median gap length (bp)     |             | 5,200       |                  | 5,728       |             | 3,581       |

# De-novo: where are we going

- Short reads assembly of maize inbred lines A632, F7, H99, HP301, Mo17, W153R
- PacBio long reads of grape cultivar Traminer (20X) and maize inbred H99 (10X)

# De-novo: where are we going



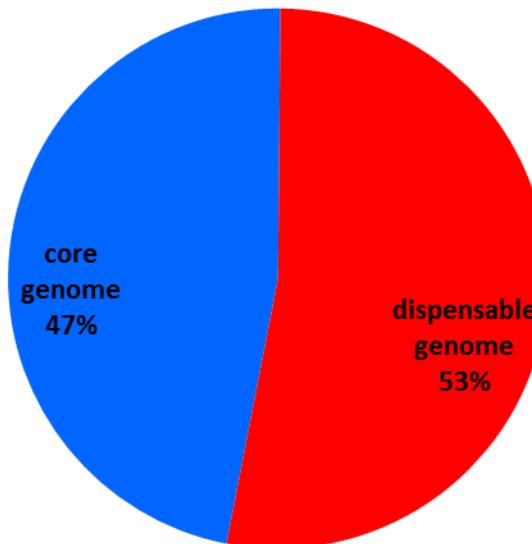
Michele Vidotto

|                       | Traminer     |                | Gouais Blanc |                | Cabernet Franc |                | Rkatsiteli   |                | Sangiovese   |                |
|-----------------------|--------------|----------------|--------------|----------------|----------------|----------------|--------------|----------------|--------------|----------------|
|                       | % of contigs | % total length | % of contigs | % total length | % of contigs   | % total length | % of contigs | % total length | % of contigs | % total length |
| <b>hemizygous</b>     | 63.55%       | 40.36%         | 49.21%       | 27.46%         | 67.81%         | 41.46%         | 60.32%       | 35.95%         | 39.75%       | 20.62%         |
| <b>non-hemizygous</b> | 36.24%       | 59.55%         | 49.44%       | 72.14%         | 31.97%         | 58.46%         | 39.37%       | 63.91%         | 60.13%       | 79.33%         |
| <b>low coverage</b>   | 0.06%        | 0.03%          | 1.26%        | 0.37%          | 0.08%          | 0.02%          | 0.15%        | 0.07%          | 0.04%        | 0.01%          |
| <b>uncovered</b>      | 0.14%        | 0.06%          | 0.09%        | 0.03%          | 0.15%          | 0.05%          | 0.16%        | 0.07%          | 0.09%        | 0.03%          |

# Size of grape pan-genome

Broad definition

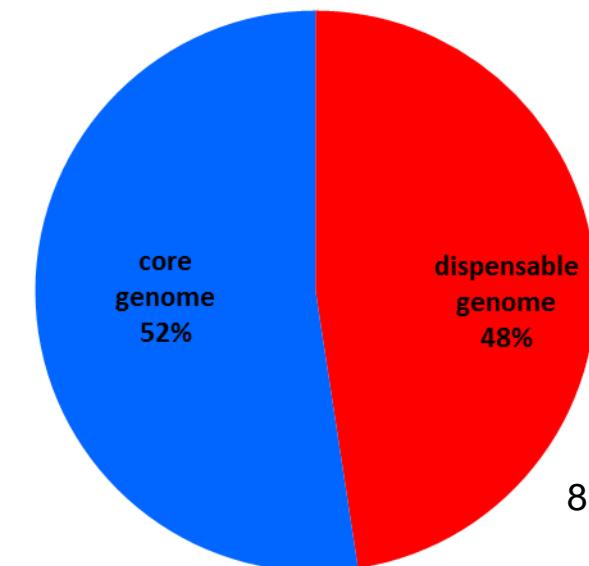
|                   | Size (Mb) |
|-------------------|-----------|
| Reference genome  | 486       |
| Pangenome         | 816.09    |
| Core genome       | 384.26    |
| Disposable genome | 431.83    |



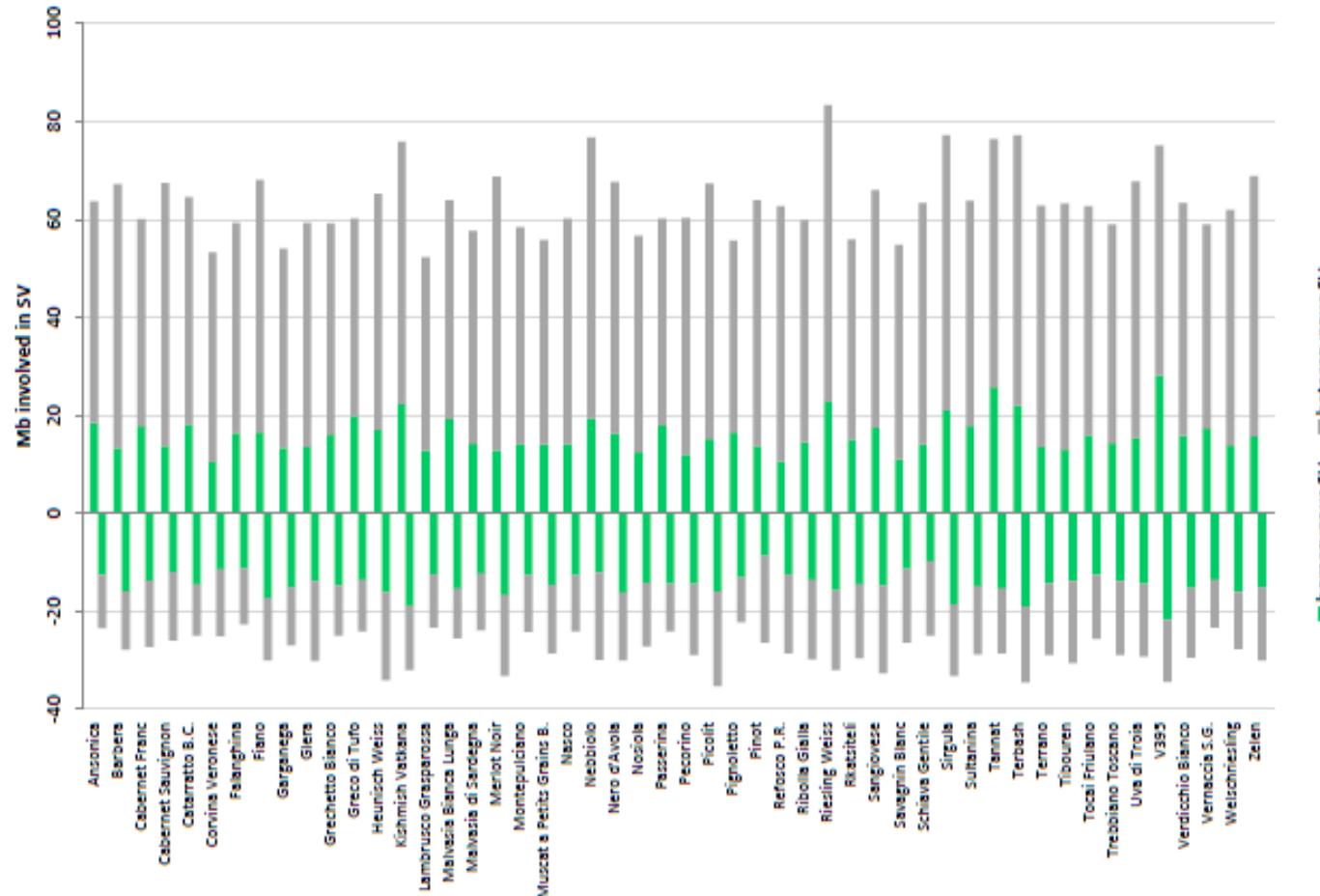
Strict definition

|                   | Size (Mb) |
|-------------------|-----------|
| Reference genome  | 486       |
| Pangenome         | 816.09    |
| Core genome       | 427.83    |
| Disposable genome | 388.26    |

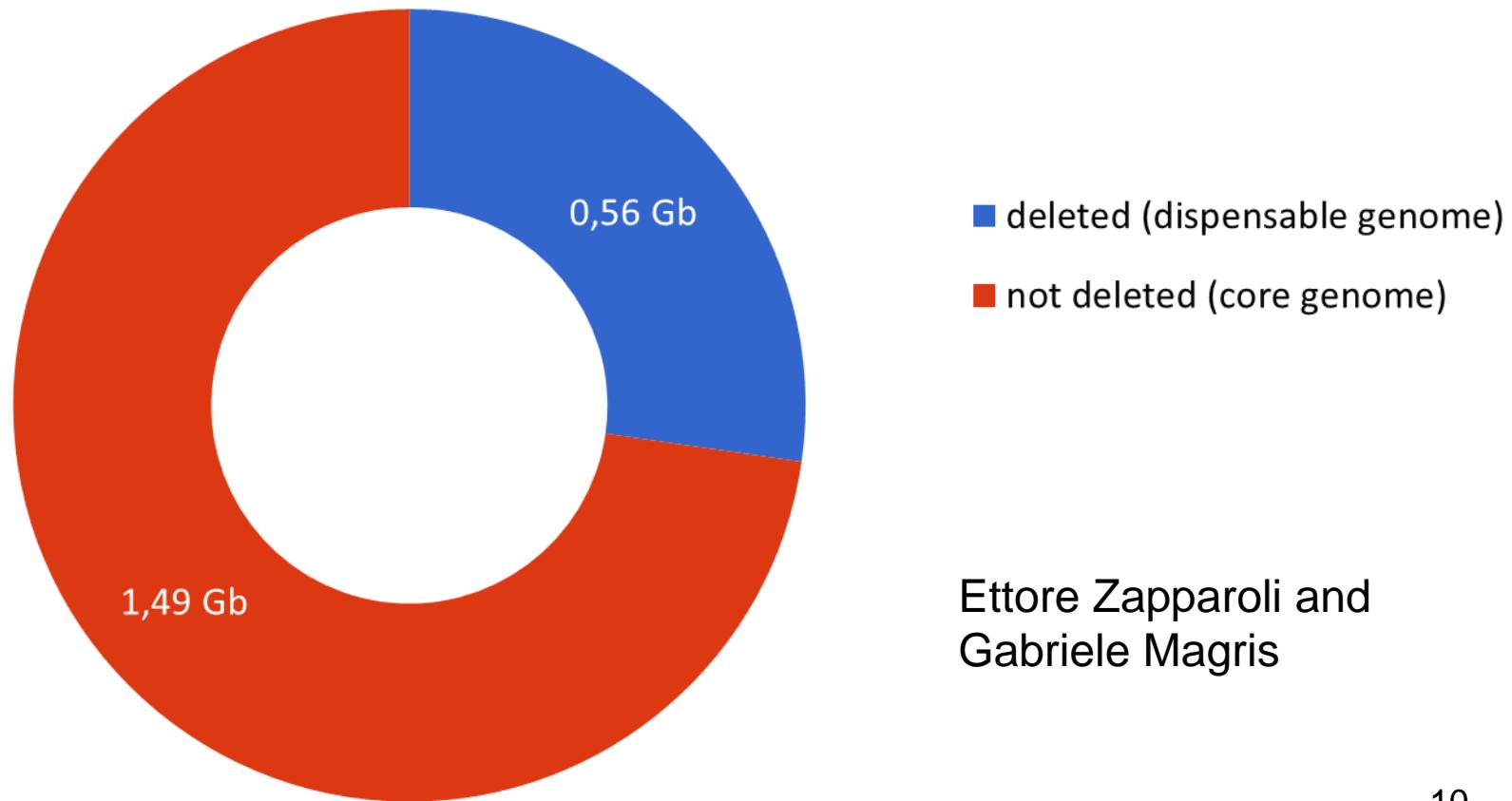
Gabriele Magris



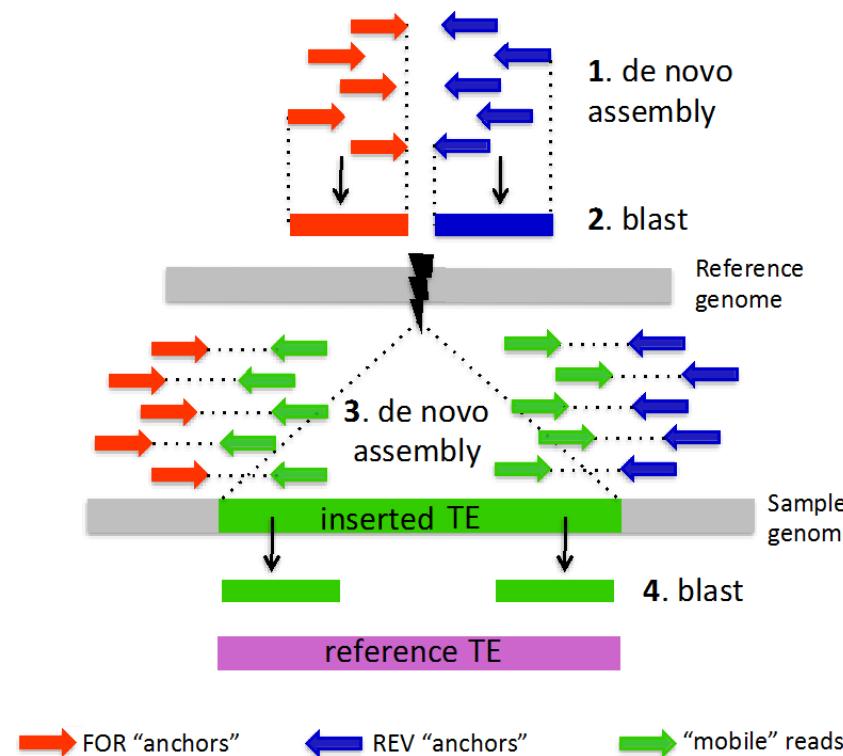
# Population-level SVs



# Preliminary results: maize pan-genome

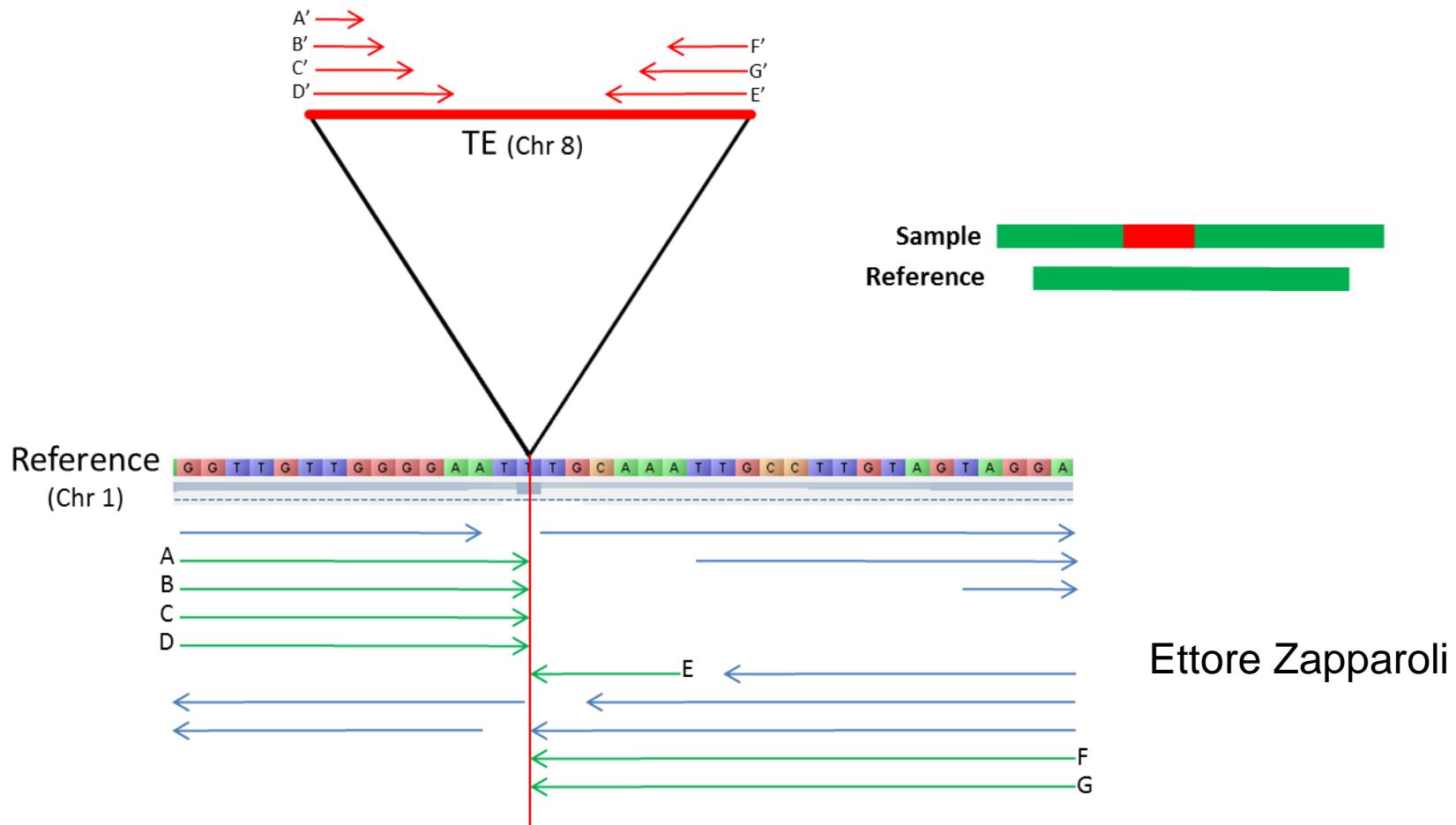


# Improved methods for SV detection: Sara's pipeline



Sara Pinosio

# Improved methods for SV detection: Wall-e



# WP1

## Expected outcomes:

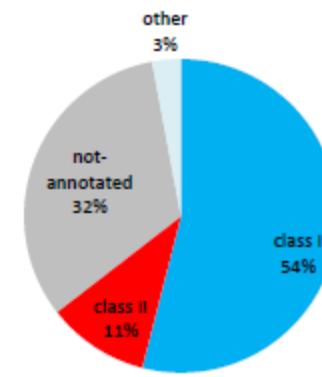
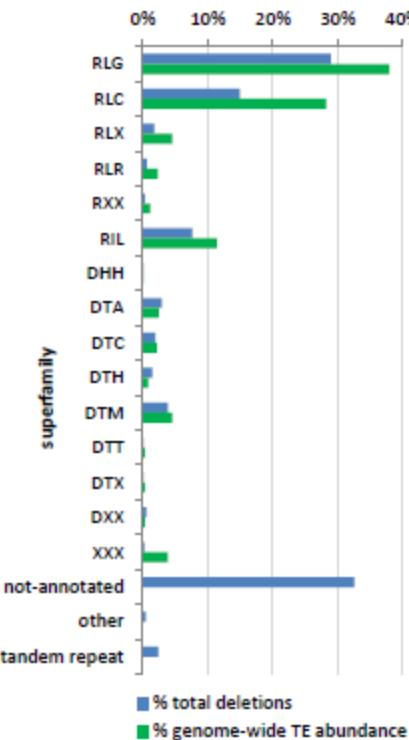
- Pan-genome reference sequences for maize and grapevine
- Pan-genome browsers for grapevine and maize
- Catalogues of SVs in maize and grapevine
- Improved methods for SV detection

# Work packages

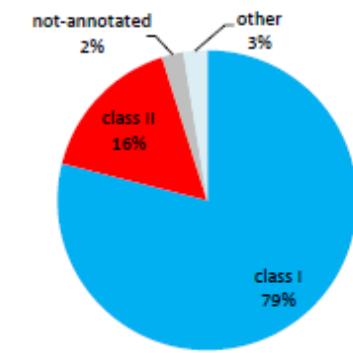
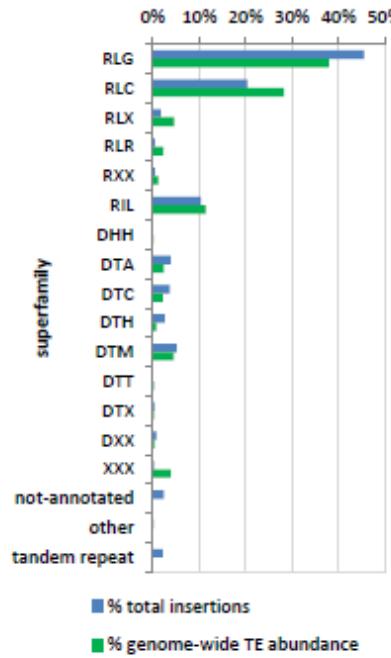
- **Workpackage 1:** Reconstructing the pan-genome in two plant species, maize and grapevine.
- **Workpackage 2:** Identifying the different mechanisms that generate and maintain the dispensable portion in the same two species.
- **Workpackage 3:** Identifying the phenotypic effects of the dispensable genome.
- **Workpackage 4:** Estimating the rates and modes of creation of new genetic variation due to dispensable genome components and its contribution to breeding progress.
- **Workpackage 5:** Extend our findings to other plant species for which the genome sequence in the meantime may have become available.

# Characterization of the pan-genome

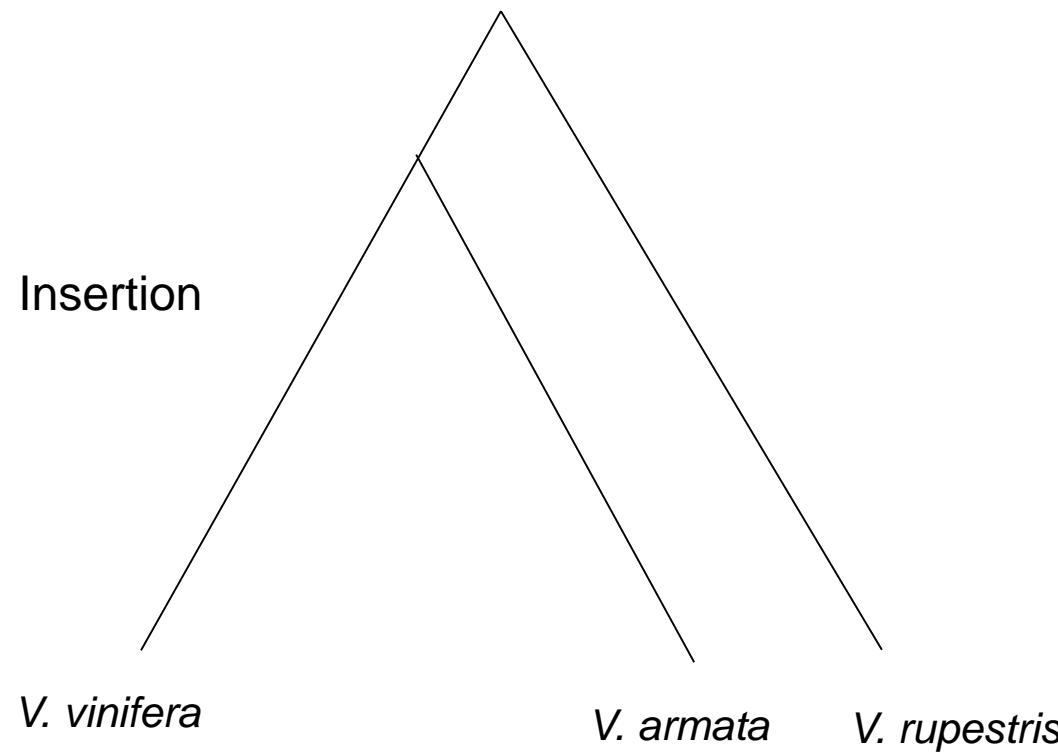
## Deletions



## Insertions



# Mechanisms of SV



*V. vinifera*

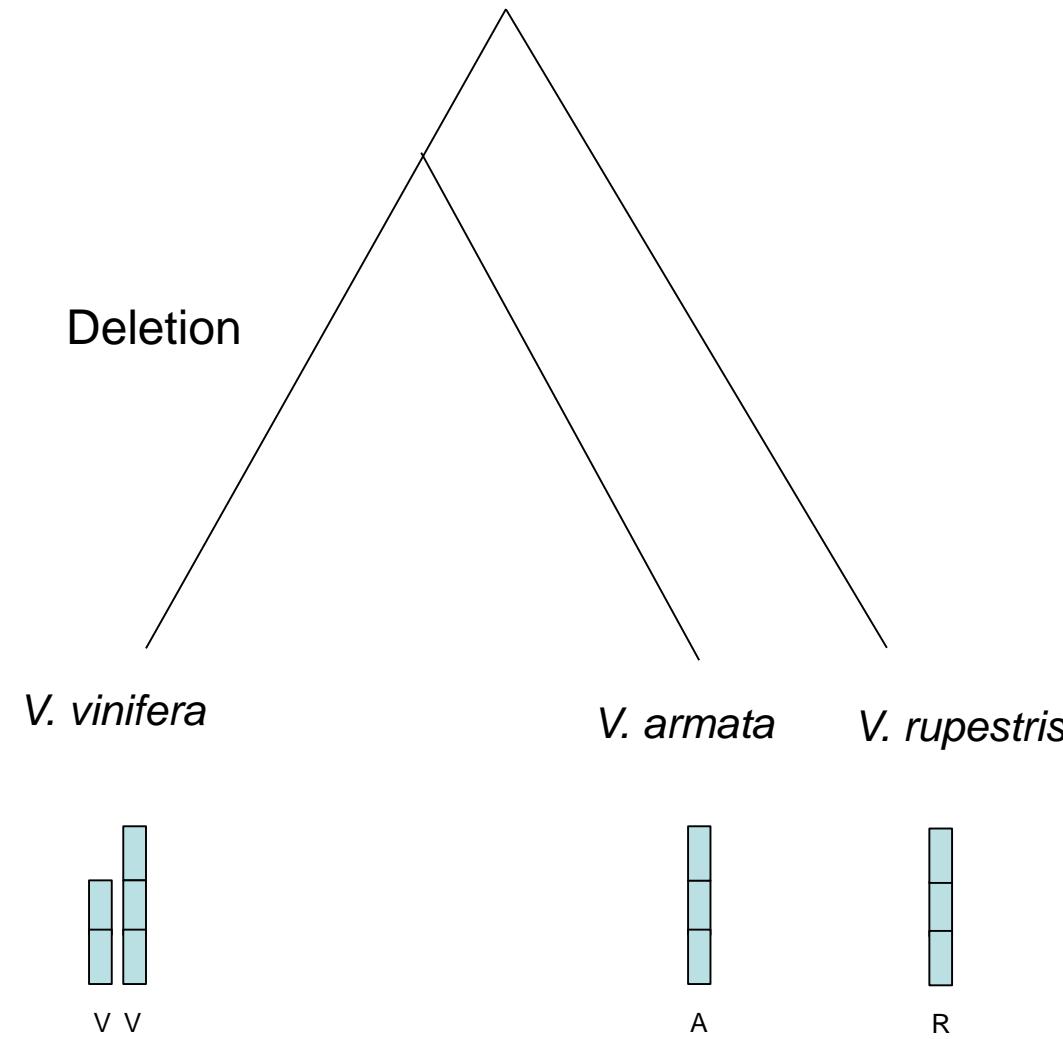


*V. armata*

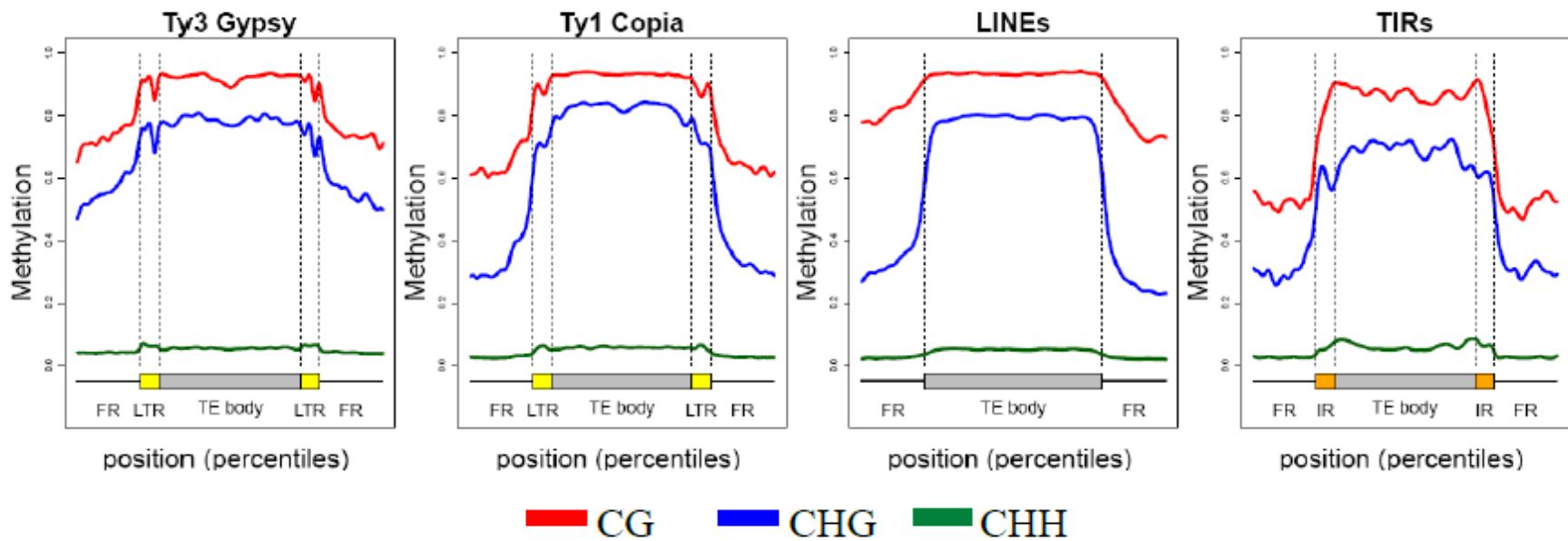


*V. rupestris*

# Mechanisms of SV



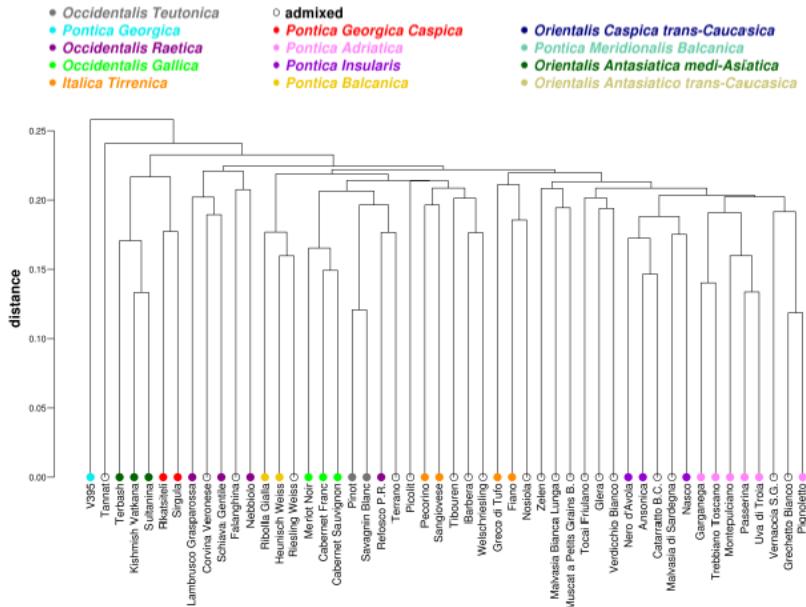
# Methylation profiles of TEs



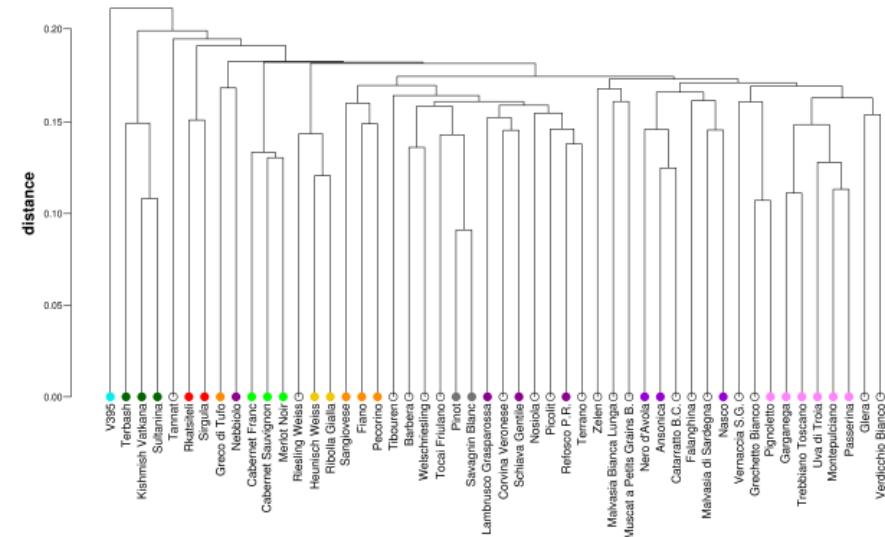
Mirko Celii e Emanuele De Paoli

# Population genomics of SVs

## Deletions

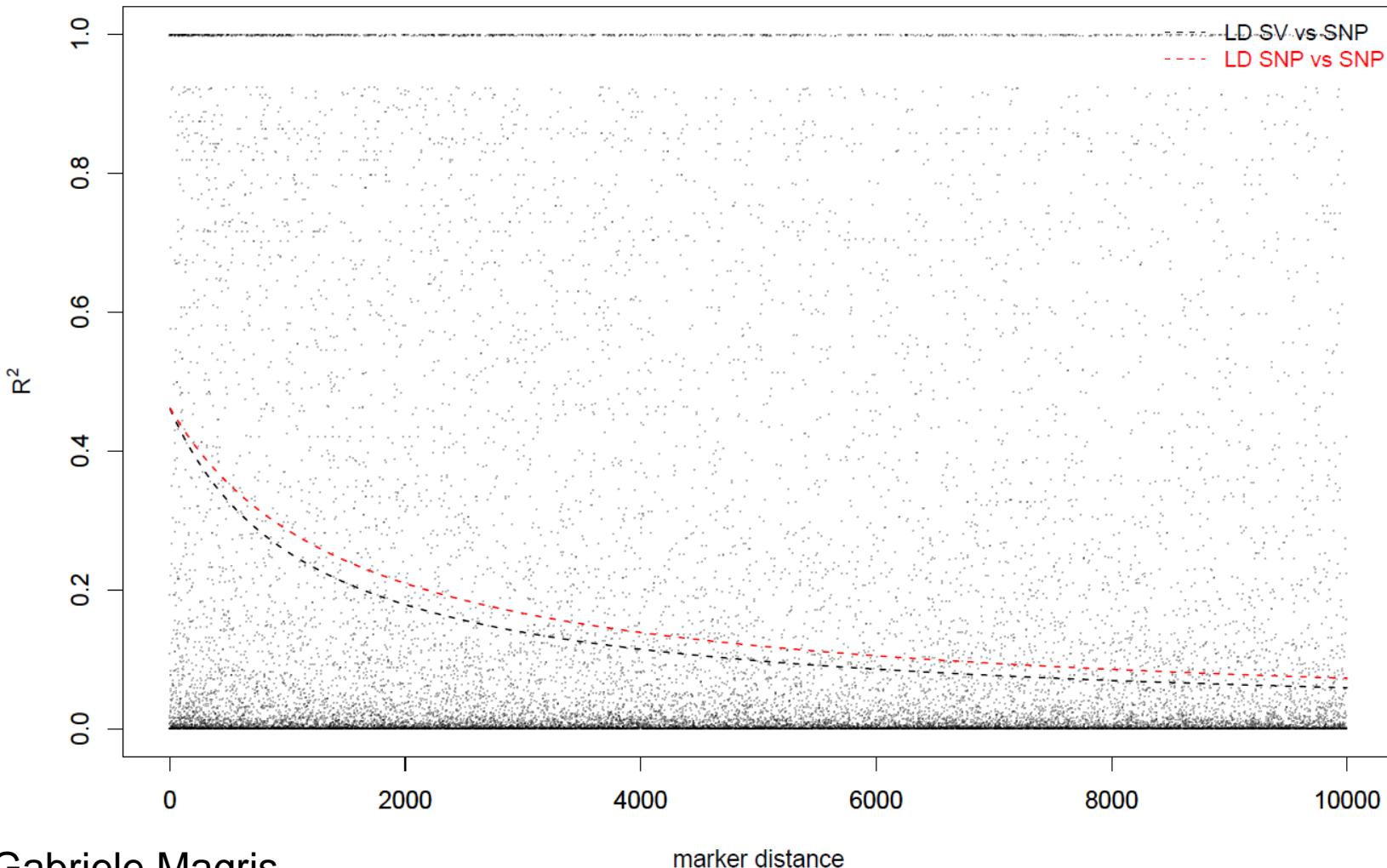


## Insertions



Gabriele Magris

# LD SNP vs SV



# WP2

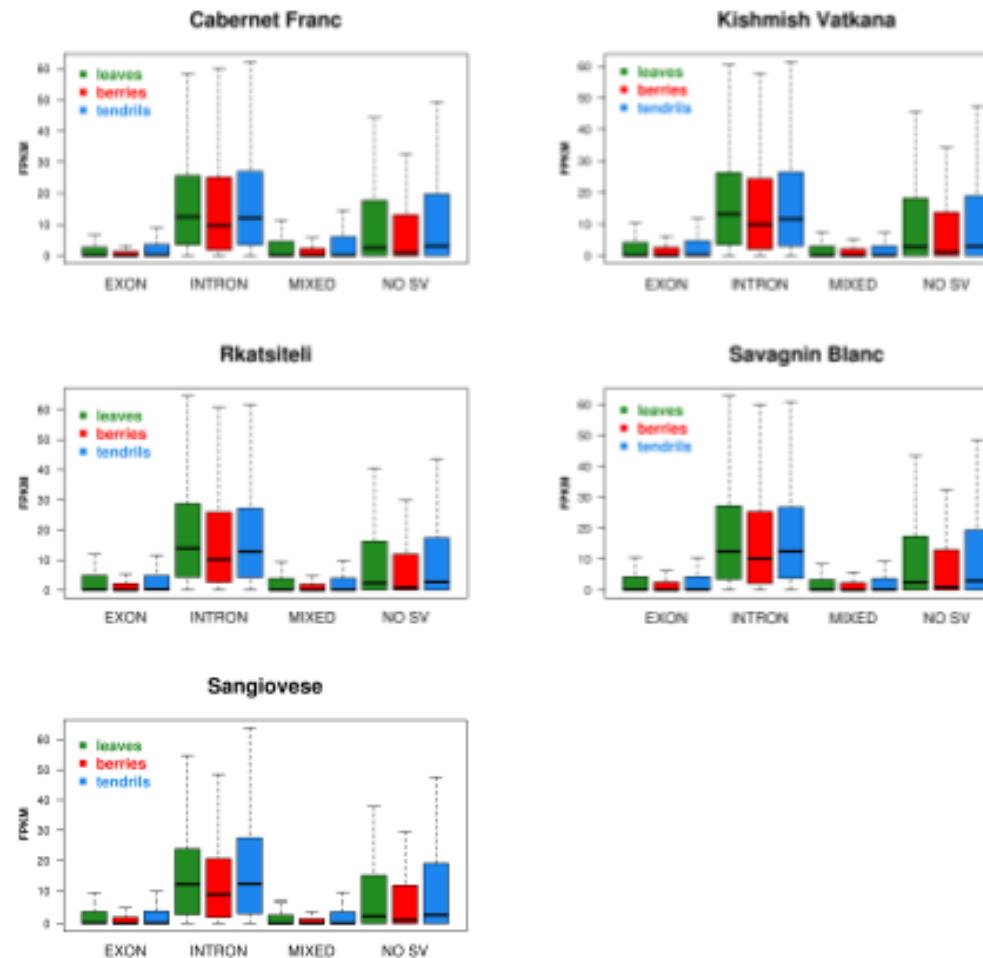
## Expected outcomes:

- Repeat and gene composition of dispensable genome
- Mutational mechanisms that create the dispensable genome
- Expression patterns and epigenetic status of dispensable genome
- Mode of evolution of SVs
- LD analysis of SVs

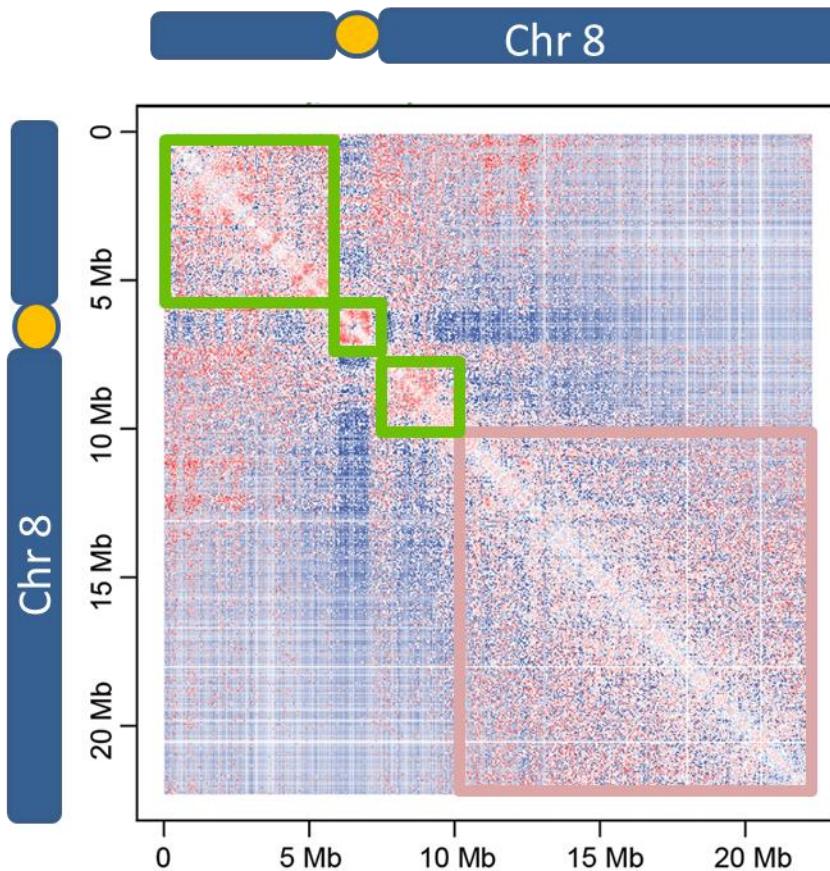
# Work packages

- Workpackage 1: Reconstructing the pan-genome in two plant species, maize and grapevine.
- Workpackage 2: Identifying the different mechanisms that generate and maintain the dispensable portion in the same two species.
- Workpackage 3: Identifying the phenotypic effects of the dispensable genome.
- Workpackage 4: Estimating the rates and modes of creation of new genetic variation due to dispensable genome components and its contribution to breeding progress.
- Workpackage 5: Extend our findings to other plant species for which the genome sequence in the meantime may have become available.

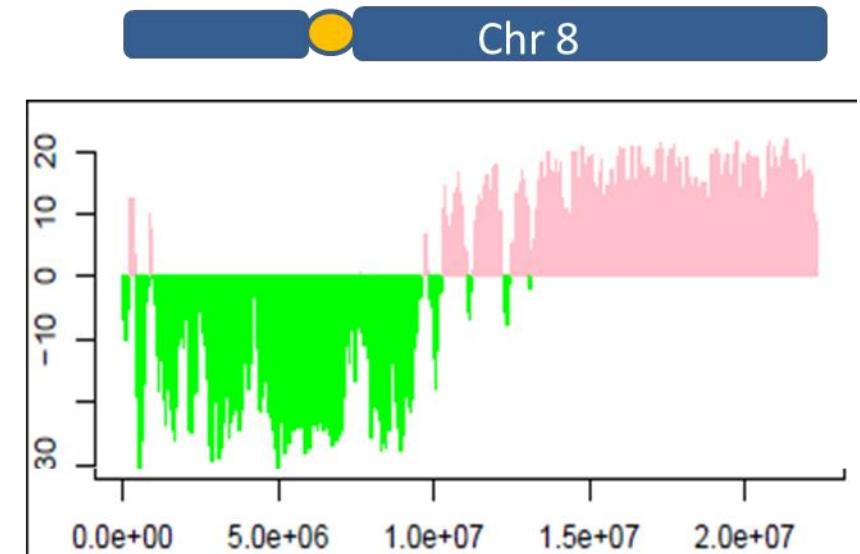
# Expression of genes affected by SV



# Hi-C



**Chr. 8 heatmap**



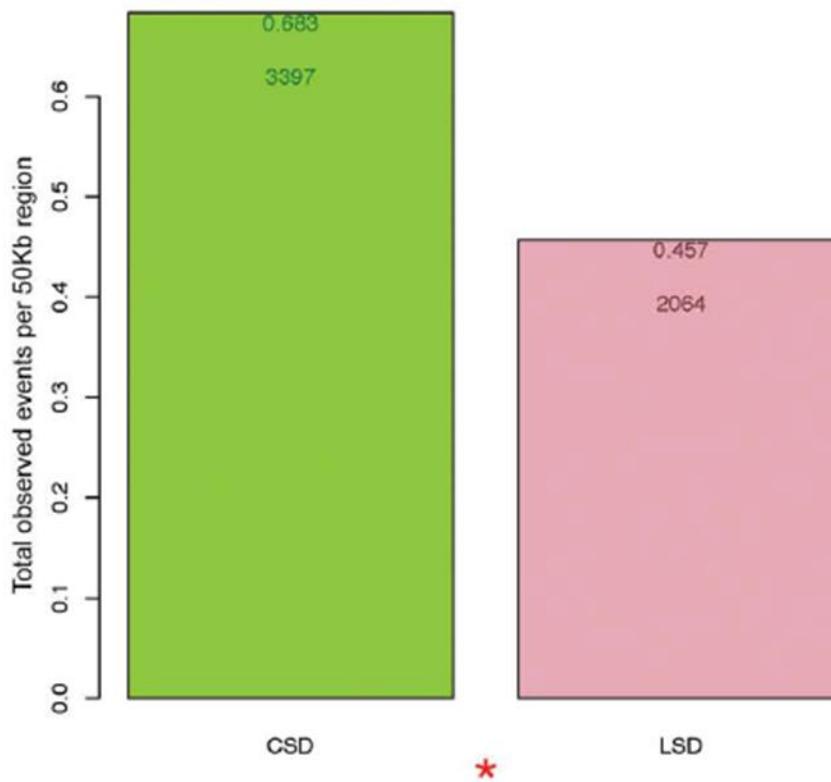
**CSD**  
Compact  
Structural  
Domain

**LSD**  
Loose  
Structural  
Domain

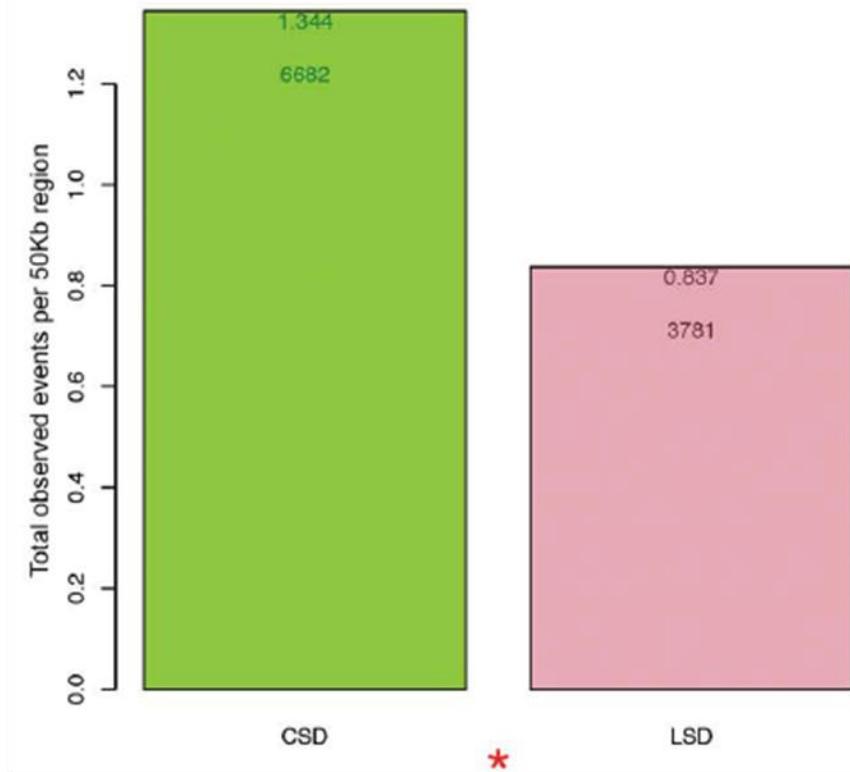
**Chr. 8 PCA**

# SV and SD

Deletions in Vitis Structural Domains



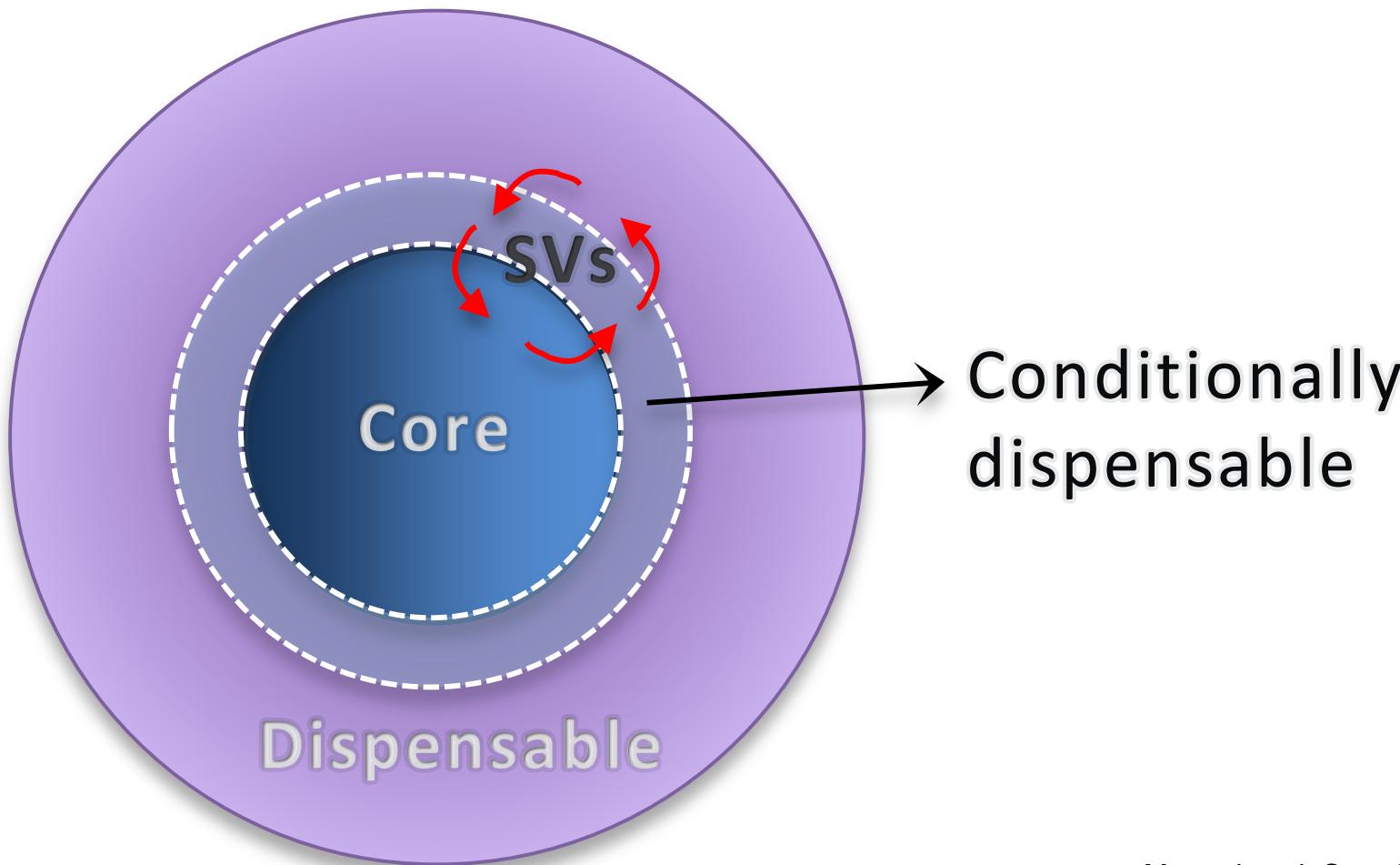
Insertions in Vitis Structural Domains



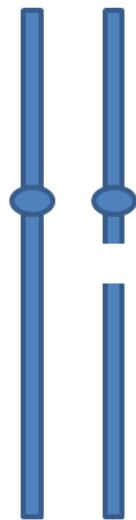
Aldo Tocci and Rachel Schwope

*Allele-specific analysis will be performed*

# Is dispensable really dispensable?



# Is dispensable really dispensable?



$A \quad a$

+ = *viable*  
- = *lethal*

Copy n.

$AA \quad + \quad 2$

$Aa \quad + \quad 1$

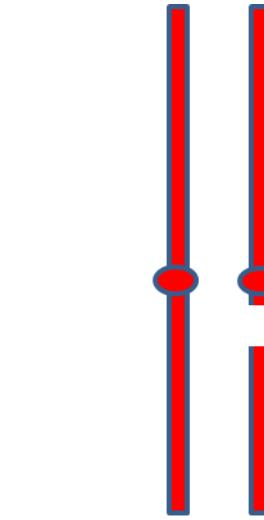
$aa \quad - \quad 0$



Not dispensable



$A \quad a$



$B \quad b$

↓  
self

$AABB \quad + \quad AaBB \quad + \quad aaBB \quad +$

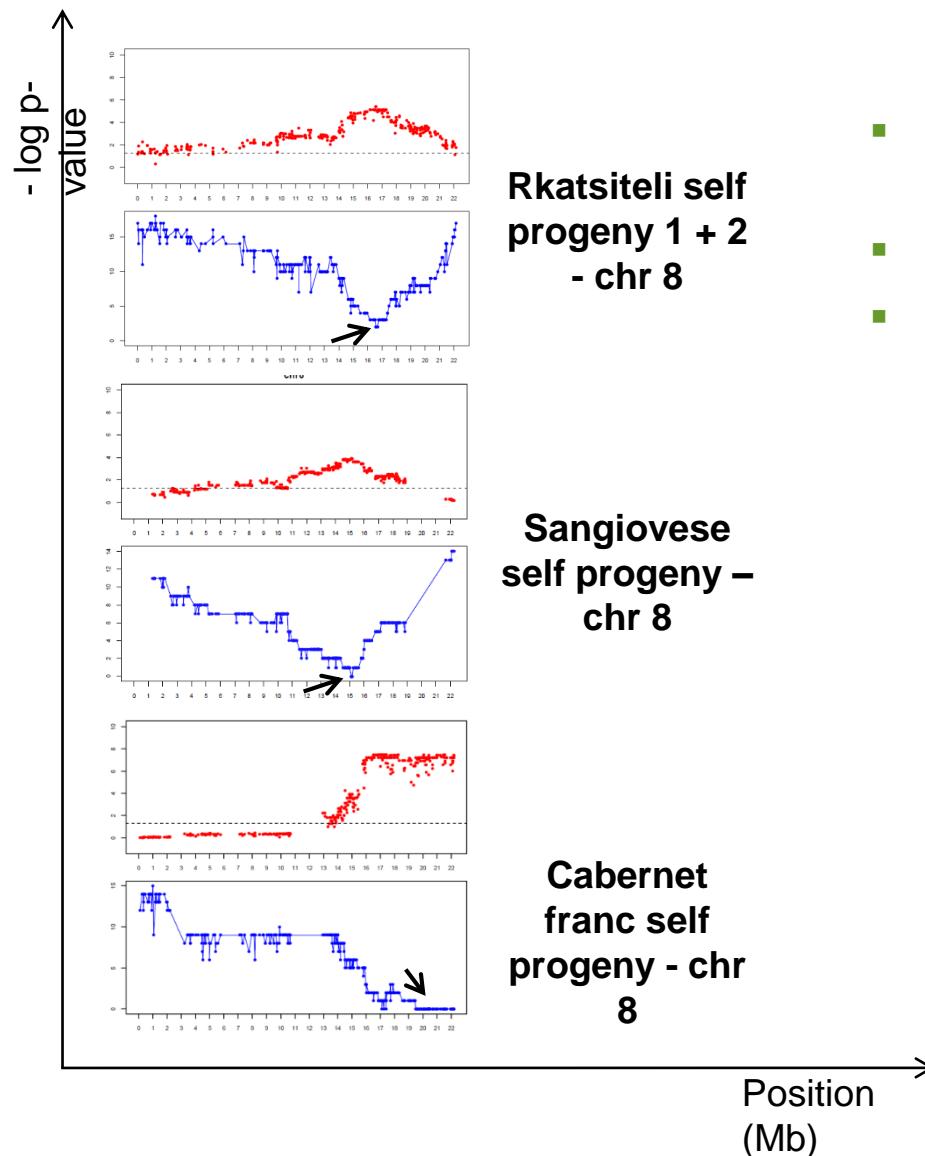
$AABb \quad + \quad AaBb \quad + \quad aaBb \quad +$

$AAAb \quad + \quad Aabb \quad + \quad aabb \quad -$



Conditionally dispensable

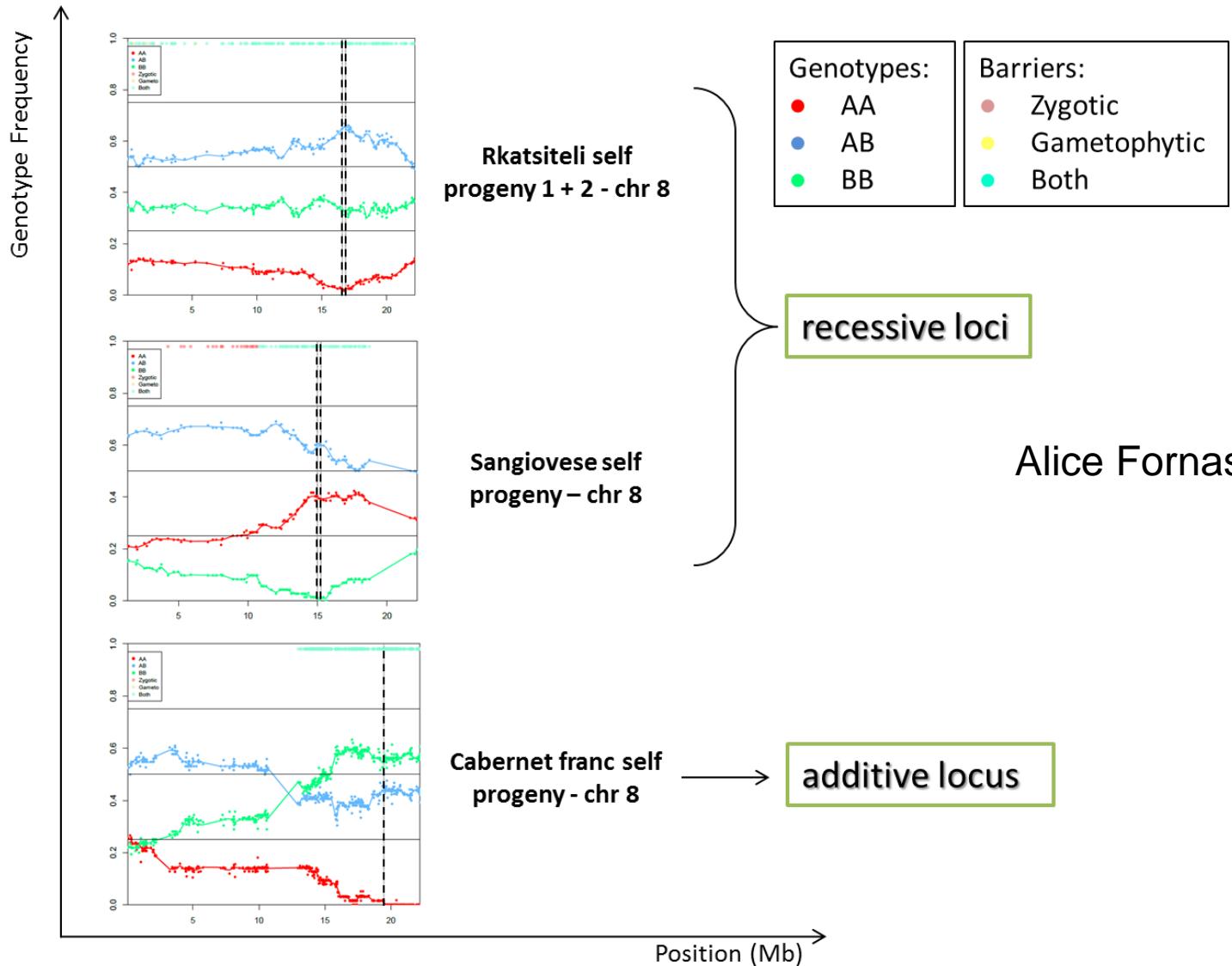
# Single locus SD in self-crosses



- Chi-square Test on single loci
- p-value of 0.05
- p-values adjusted for *fdr control*

Alice Fornasiero

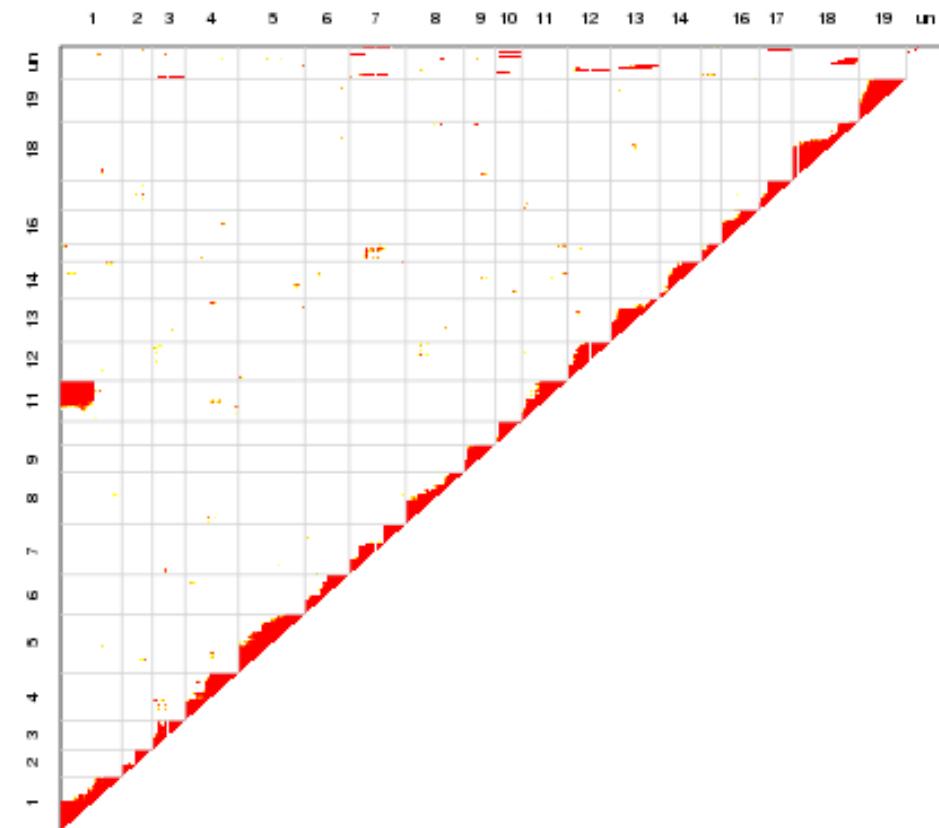
# Single locus SD in self-crosses



# Two-loci SD in selfed Rkatsiteli

Segregation distortion  
when considering regions  
of chr 1 and chr 11 that  
show strong association:  
pseudolinkage

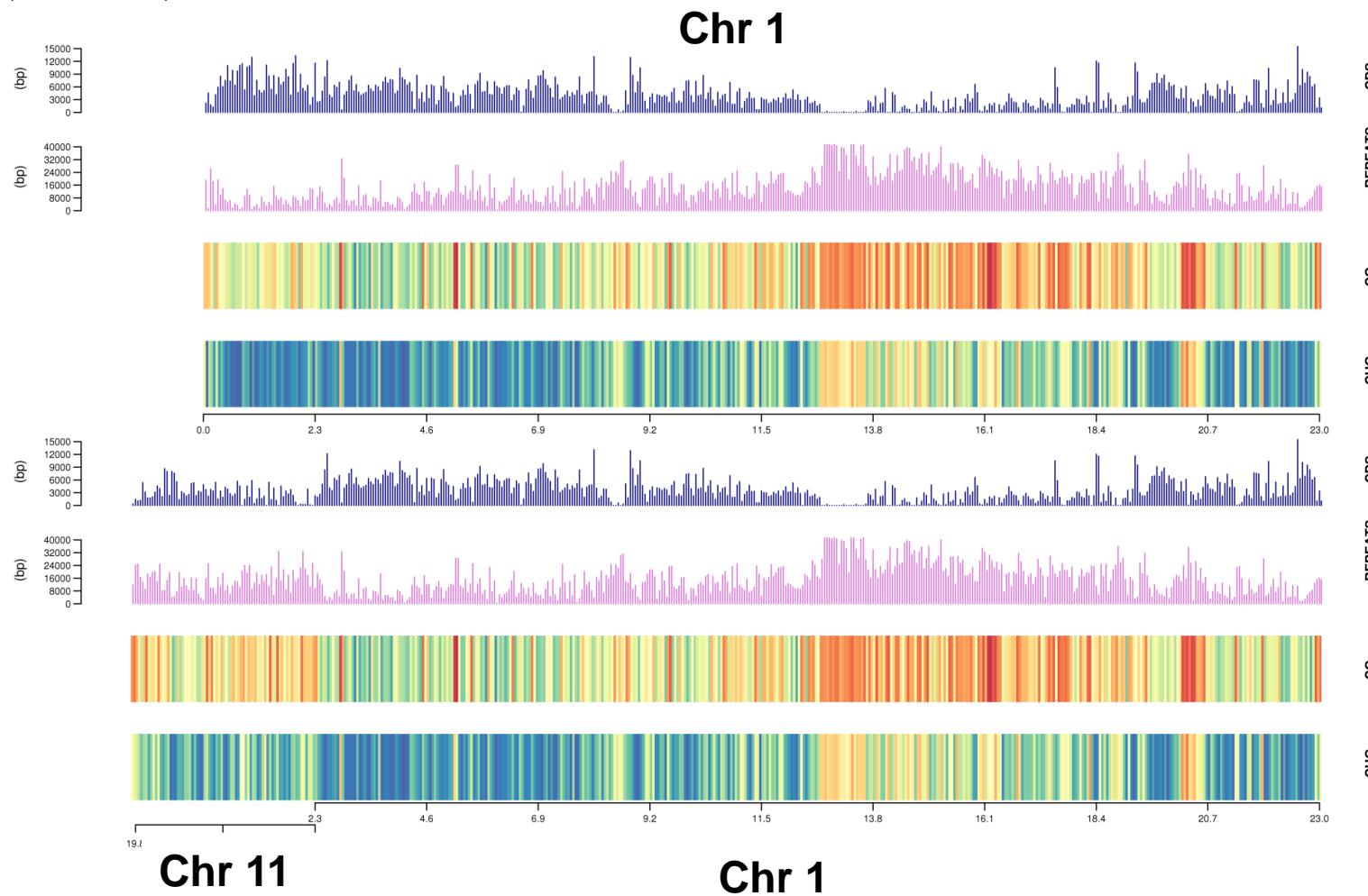
- **excess** of some double-homozygous (**AABB**, **aabb**) and double-heterozygous (**AaBb**)
- **lack** of all other 6 genotypic classes (**aaBB**, **AAbb**, **Aabb**, **AaBB**, **AABb**, **aaBb**)



Alice Fornasiero

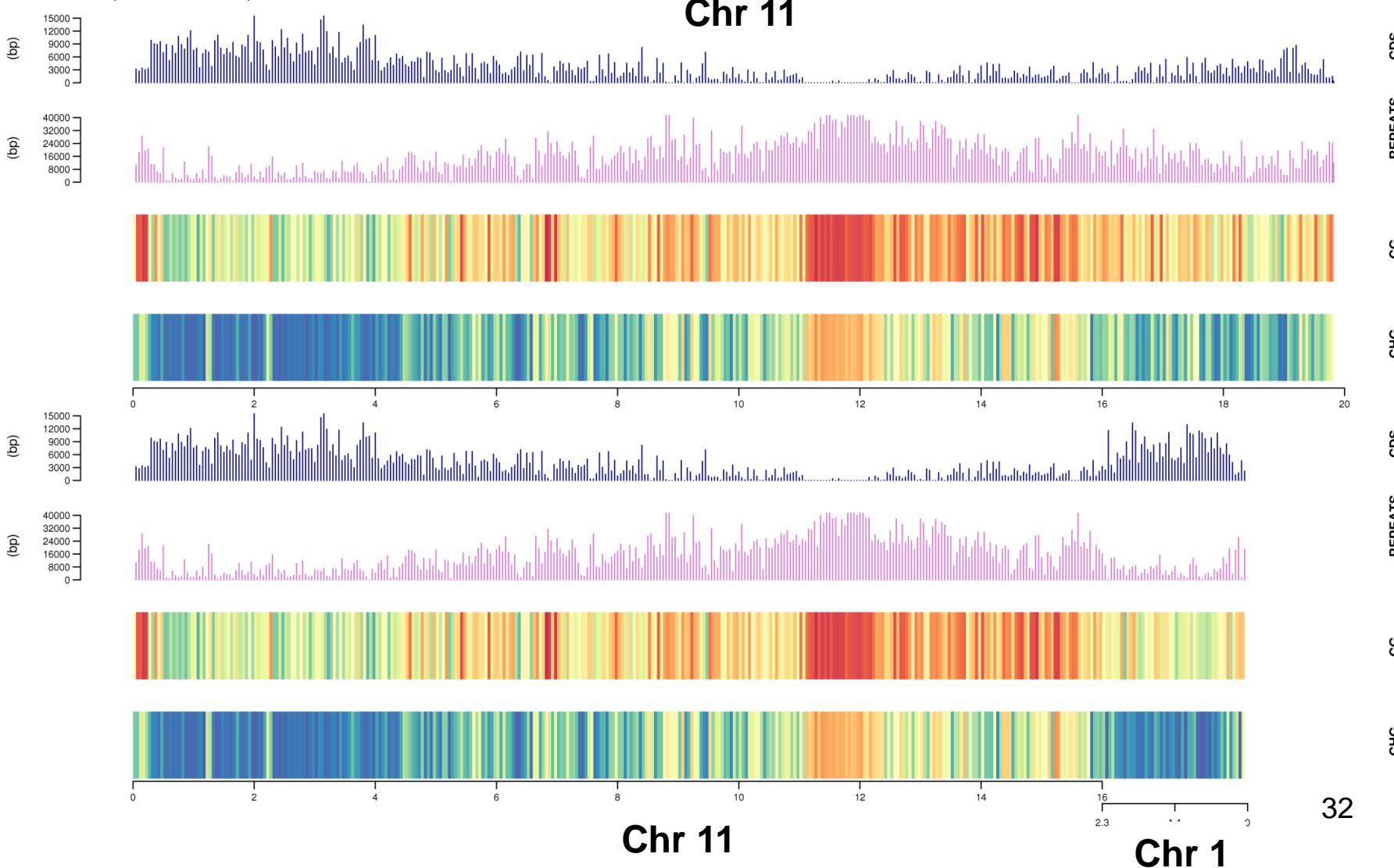
# Chromosomal structure with and without translocation

Alice, Gabbo, Mirko

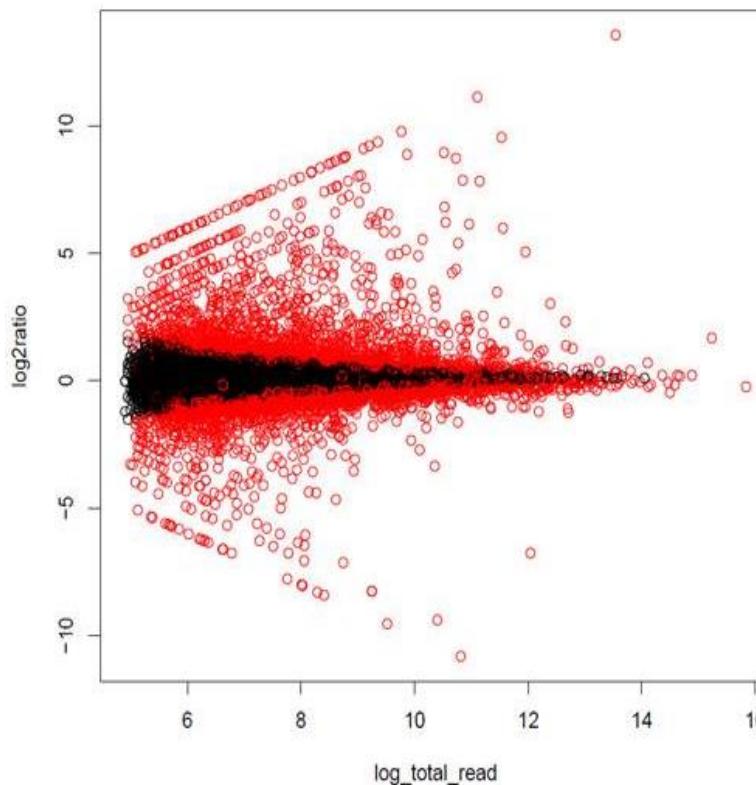
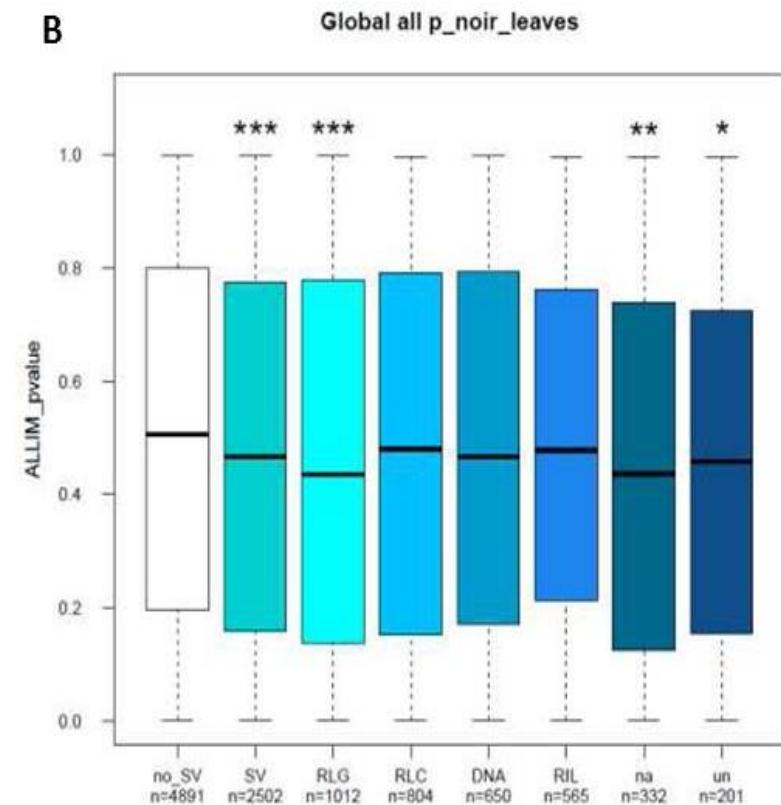


# Chromosomal structure with and without translocation

Alice, Gabbo, Mirko

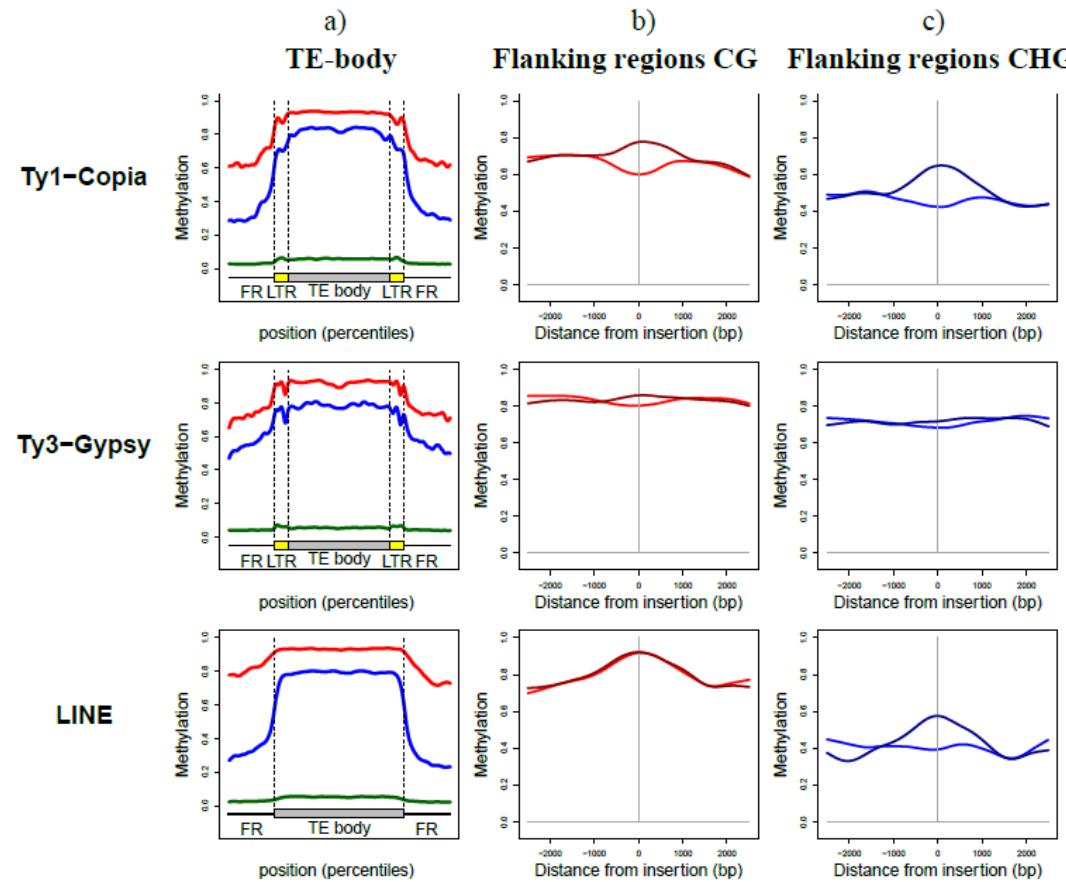


# SV and ASE

**A****B**

Eleonora Paparelli

# SV and ASM



Mirko Celii and Emanuele De Paoli

# WP3

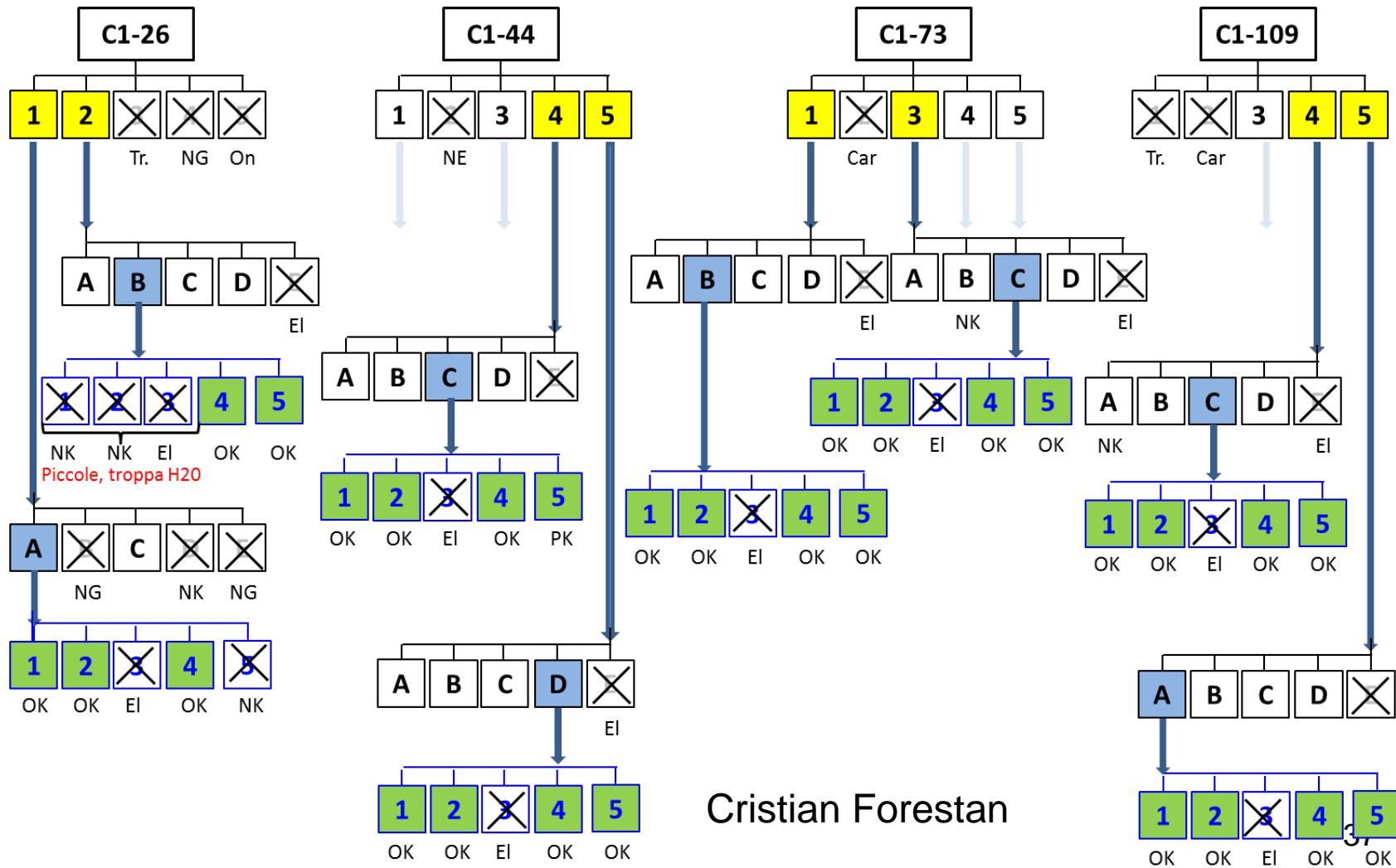
## Expected outcomes

- Correlation analysis between cis-regulatory variation and TE-related SVs
- Regulatory effects of genic CNVs
- Effects of large SVs on heterosis (maize) and disease resistance (grapevine)

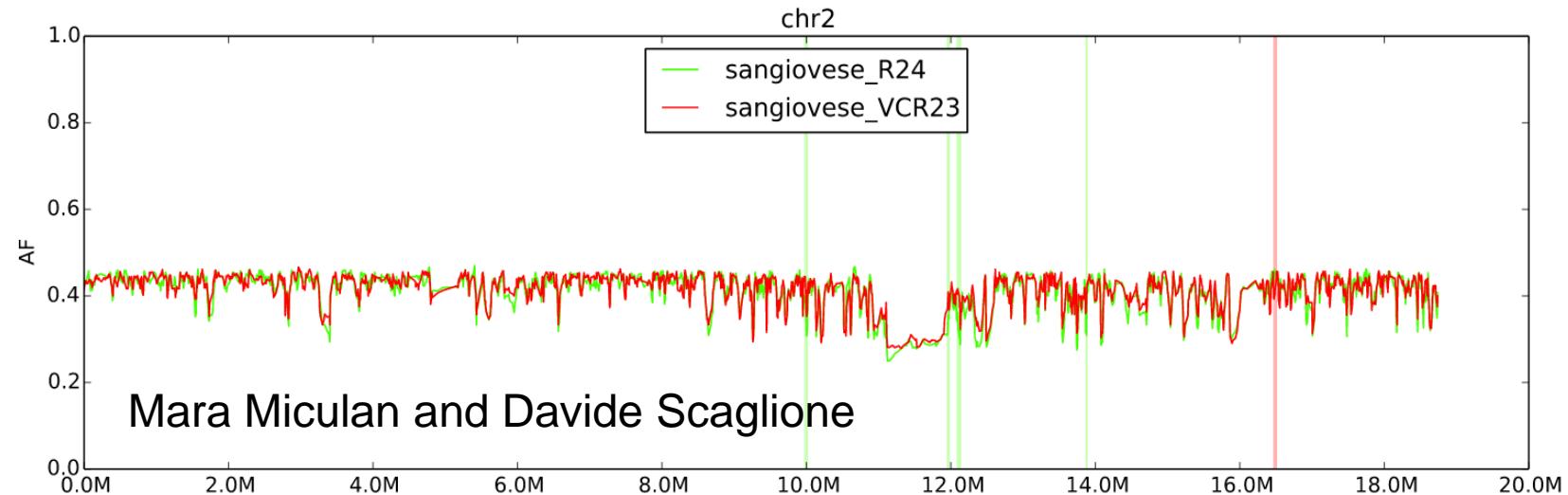
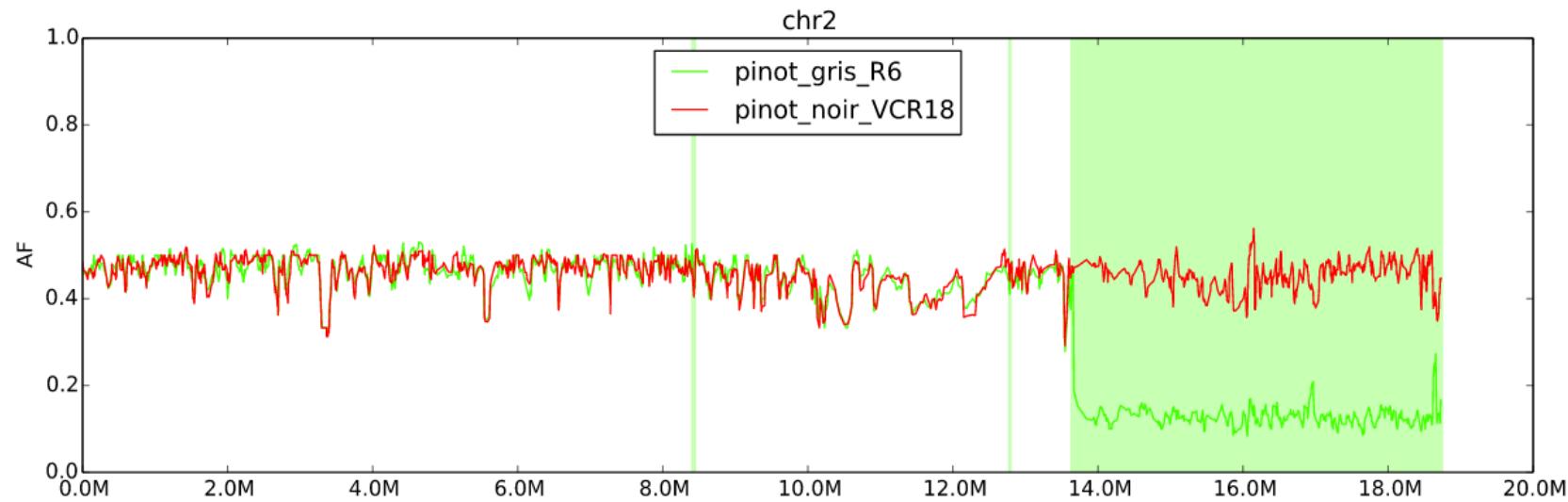
# Work packages

- Workpackage 1: Reconstructing the pan-genome in two plant species, maize and grapevine.
- Workpackage 2: Identifying the different mechanisms that generate and maintain the dispensable portion in the same two species.
- Workpackage 3: Identifying the phenotypic effects of the dispensable genome.
- Workpackage 4: Estimating the rates and modes of creation of new genetic variation due to dispensable genome components and its contribution to breeding progress.
- Workpackage 5: Extend our findings to other plant species for which the genome sequence in the meantime may have become available.

# Novel SV in maize



# Somatic SV in grape



# Induzione di SV: HSP90 and PIN1

- Inibizione di HSP90 correlata con aumento di mobilità TE
- Silenziamento di PIN1 (in *A. thaliana*) correlato con aumento di mobilità TE
- Work in progress

# WP4

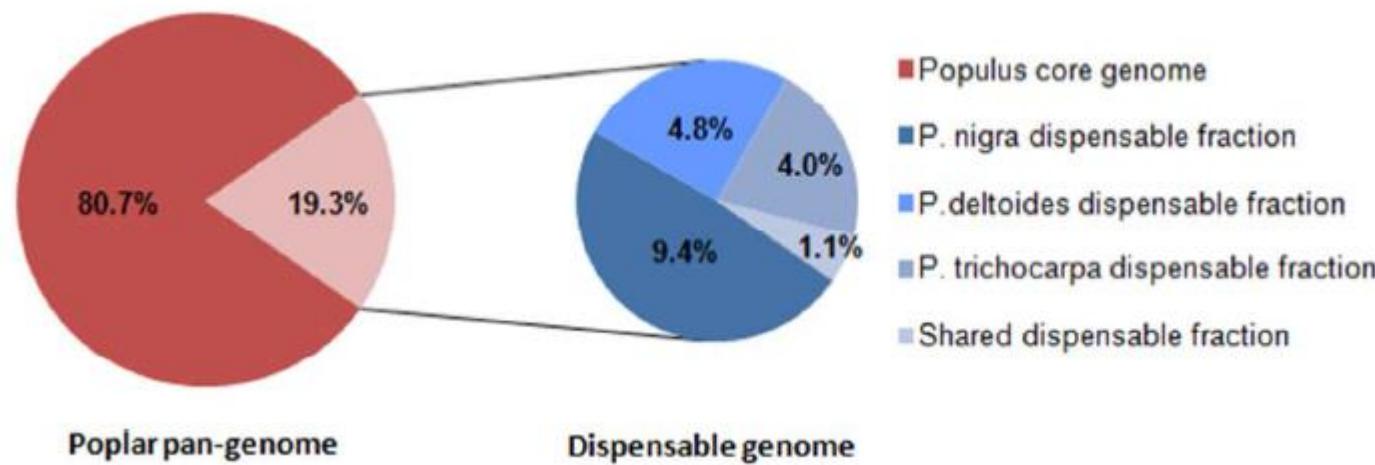
## Expected outcomes

- Somatic *de novo* mutation rate for SV in grape: relative effect of outer and inner tunica
- Germline *de novo* mutation rate for SV in maize
- Correlation between inhibition of Hsp90 and transcription levels of TEs
- Induction of new structural variants by inhibition of the host genome control

# Work packages

- **Workpackage 1:** Reconstructing the pan-genome in two plant species, maize and grapevine.
- **Workpackage 2:** Identifying the different mechanisms that generate and maintain the dispensable portion in the same two species.
- **Workpackage 3:** Identifying the phenotypic effects of the dispensable genome.
- **Workpackage 4:** Estimating the rates and modes of creation of new genetic variation due to dispensable genome components and its contribution to breeding progress.
- **Workpackage 5:** Extend our findings to other plant species for which the genome sequence in the meantime may have become available.

# Poplar



Sara Pinosio (submitted for publication)

# WP5

## Expected outcomes

- Identification of somatic SVs in the *Citrus* genus
- Identification of SVs in peach

# Acknowledgments

All Novabreed, IGA and IGATS staff



European Research Council

Established by the European Commission

# Acknowledgments

All Novabreed, IGA and IGATS staff  
And Michele Morgante for winning the grant!



European Research Council

Established by the European Commission