The NOVABREED project

The plant Pan-Genome consists of:

- Core genomic features
- Dispensable Genome due to Structural Variants (SVs)

(Marroni, Pinosio and Morgante, 2014)

- Uncovering the composition, origin and structure of the SVs
- Understand the contribution of the SVs to the creation of new genetic variation in plant
Structural variants (SVs)
(Marroni, Pinosio and Morgante, 2014)

- **CNV**: Copy-Number Variants

  ![Sample and Reference Comparison]

- **PAV**: Presence-Absence Variants
  
  » deletions and insertions
  
  (TEs movement and genomic rearrangements)

  ![Sample, Reference, and Variants Comparison]
Aim

Assess the effect of the SVs on *Vitis vinifera* genome architecture

a. Investigate the role of genomic architecture in the interplay between structure, sequence and function

b. Investigate allele-specific chromatin structure:
   • its dependence on SVs in haplotypes
   • its effects on allele-specific regulation

c. Assess variation of genome architecture in different *Vitis vinifera* varieties
The standard Hi-C Method

*V. vinifera* young leaf

(Lieberman-Aiden, *et al.* 2009)
(van Berkum, *et al.* 2010)
What has been done…

Variety: Pinot Noir

1. Contact maps reconstruction via Hi-C

2. Global Structural Domains (SDs) conformation analysis

3. Hi-C data for scaffolding of an almost complete *Vitis vinifera* assembly
1. Contact Maps Reconstruction

Hi-C reads are aligned on current *Vitis vinifera* reference genome

**Resolution**: 1 Mb

**Red** = high interaction frequency

**Blue** = low interaction frequency

(Heinz, *et al.* 2010)
1. Contact Maps Reconstruction

There are Structural Domains (SDs) inside chromosomes

Resolution: 50 Kb

Red = high interaction frequency

Blue = low interaction frequency

(Heinz, et al. 2010)
2. Global Structural Domains (SDs) conformation analysis via PCA

**Principal Component Analysis (PCA)** simplifies the data into 2 sets of interactions: sparse and condensed, identifying two different SDs.

**CSD**
(Compacted Structural Domain): inactive epigenetic marks, low gene expression, presence of TEs and small RNAs.

**LSD**
(Loose Structural Domain): active histone modifications, high transcription levels.

(Grob, Schmid and Grossniklaus, 2014)
2.a Correlate SDs with the genomic features of Pinot noir

CG_methylation in Vitis Structural Domains

expression (log10) in vits SDs
2.a Correlate SDs with the genomic features of Pinot noir

**Whole Genome Number of Genes**
- CSD: 2,541
- LSD: 22,098

**H3K4me3 Levels across Vitis SDs**
- CSD: 0.709
- LSD: 2.653

* indicates significant difference.
2.a Correlate SDs with the genomic features of Pinot noir
3. Improve assembly using Hi-C interaction data

**chrUn**: set of scaffolds that could not be associated to any chromosome during the assembly.
3. Improve assembly using Hi-C interaction data

LACHESIS
(Ligating Adjacent Chromatin Enables Scaffolding In Situ): a computational method that exploits the genomic proximity signal in Hi-C data sets for ultra-long range scaffolding of de novo genome assemblies.

IMPORTANT: it doesn’t require any reference genome, only Hi-C data!

(Burton, et al. 2013)
3. Improve assembly using Hi-C interaction data

Clustering improvement:

Total scaffolds: 2059

chrUn scaffolds: 1849/2059

Assigned chrUn scaffolds: 1834/1849

~ 39 Mb added
3. Improve assembly using Hi-C interaction data
what is going on...

Variety: Rkatsiteli

1. Contact map reconstruction via *in situ* Hi-C

2. Lachesis for *de novo* assembly scaffolding, improving N50 and L50

3. Haplotype-specific Hi-C
1. *In Situ* Hi-C

What’s different from classic Hi-C?

The DNA-DNA proximity ligation process happens **inside** the intact **nuclei** of permeabilized crosslinked cells.

Advantages:

- Reduced frequency of spurious contacts due to random ligation in diluted solution
- Faster protocol (requiring 3 days instead of 7)
- Enables higher resolution (up to ~1Kb) 

(Rao, *et al.* 2014)
2. Improvement of \textit{de novo} assembly

Rkatsiteli \textit{de novo} assembly summary

Estimated genome size: 486,2 Mbp

Number of scaffolds: 10,089

\textbf{L50} scaffold length: 352,572 bp

\textbf{N50} scaffold count: 612

(M. Vidotto, 2015)
3. Obtain allele-specific versions of Hi-C maps

Rkatsiteli (heterozygous)

Haplotype ♀

Haplotype ♂

- gene
- class I TE
- class II TE
Future Perspectives

1. Finalize the ongoing works

2. Obtain high-resolution data from *in situ* Hi-C to identify promoter-enhancer interactions
Acknowledgments

Prof. Michele Morgante

Fabio Marroni
Rachel Schwope

Simone Scalabrin
Mirko Celii
Alice Fornasiero
Gabriele Magris
Ettore Zapparoli

Everyone from IGA (Istituto di Genomica Applicata) and Di4a (Dipartimento di Scienze AgroAlimentari, Ambientali e Animali - UNIUD)
THANKS and Hi-C YOU SOON!