Finding the Driving Forces of Recombination Frequency in *Vitis vinifera*

Alice Fornasiero [1,2,*], Aldo Tocci [1,2,3,*], Rachel Schwope [1,2], Fabio Marroni [1,2], Gabriele di Gaspero [2], Michele Morgante [1,2]

[1] Dipartimento di Scienze Agroalimentari, Ambientali e Animali, Università di Udine, Udine, Italy - [2] Istituto di Genomica Applicata, Udine, Italy
[3] Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy

*** Both authors contributed equally to this work.**

e-mail: **afornasiero@appliedgenomics.org**

Introduction

Marked differences in recombination per unit of physical distance along chromosomal domains have been reported in several crops species.
Recombination frequency might be affected by genomic properties thus reflecting functional properties.
We set out to **characterize genome-wide recombination rate** in *Vitis vinifera* at three different levels of analysis: from **sequence features**, such as insertion and deletion of transposable elements (TEs), genes and repetitive DNA; to **epigenetic modifications**, such as DNA and histone methylation; up to **3D features**, like chromatin conformation and accessibility.
This work aimed to reconstruct the panorama of changes in the physical structure of chromosomes that could affect crossover occurrence.

Methods

Genotyping-by-Sequencing data in self-crosses from three different varieties (Pinot Noir, Schiava Grossa and Rkatsiteli) were used to generate genetic maps and to estimate recombination frequency in grapevine.

Several chromatin features were studied in the Pinot Noir variety, by means of:

- **Hi-C method** for chromatin conformation analysis;
- **ATAC-Seq** for detection of accessible chromatin;
- **Chip-Seq** to examine histone methylation associated with active transcription of genes ($H3K4^{me3}$);
- **BS-Seq** to analyse DNA methylation in CG and CHG contexts.

Hi-C data on chromosome contact was used to compute the ratio of inter- to intra- chromosomal interactions (**RI value**).

Results

Characteristics of genomic regions along chromosome 8

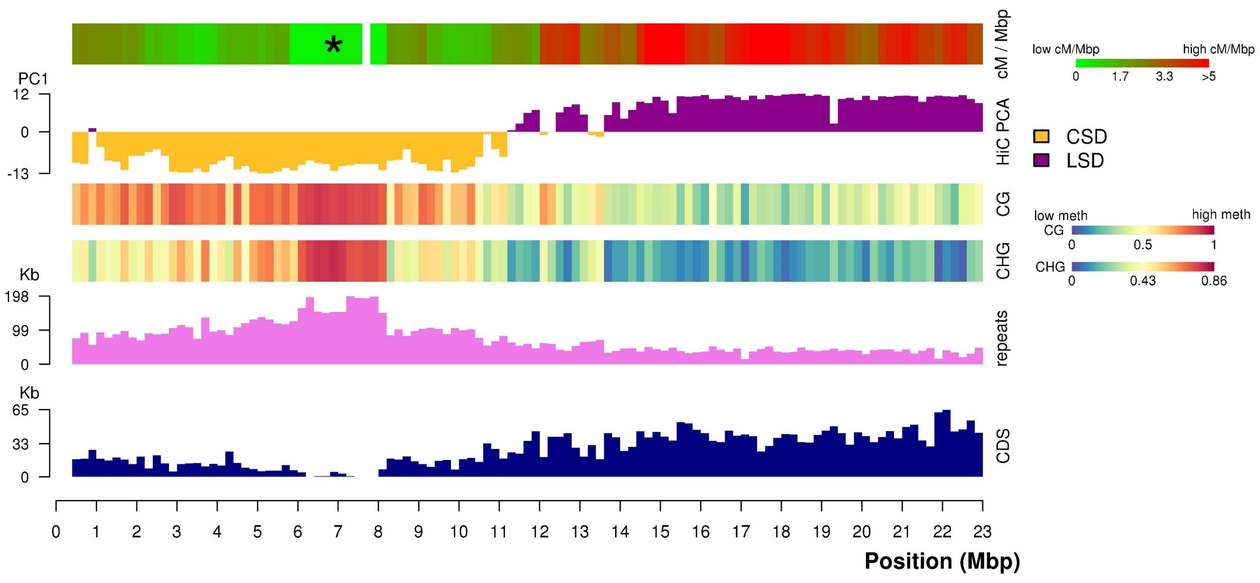

Figure 1-A. *From top to bottom: recombination frequency (cM/Mb); principal component analysis (PCA) of Hi-C data defining regions of open chromatin (LSD, violet) and closed chromatin (CSD, yellow); methylation in CG and CHG contexts; density of repetitive DNA (repeats, pink) and coding sequence (CDS, blue). Features were calculated in bins of 200Kb. *: centrosome.*

High-recombining regions:

- located in a open chromatin context (loose structural domain, LSD);
- showed high levels of gene density and low levels of repetitive DNA and DNA methylation;
- characterized by a genomic scenario reflecting actively transcribed regions.

Low-recombining regions:

- located in a closed chromatin context (compact structural domains, CSD);
- presented high levels of repetitive DNA and DNA methylation and low levels of gene density;
- showed characteristics coherent with an inactive transcriptomic context (Fig. 1A-B).

Recombination and chromatin conformation

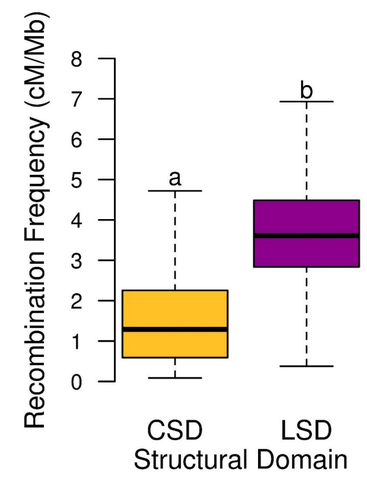

Figure 1-B. *Genome-wide distribution of recombination frequency within structural domains was tested with Mann-Whitney U test (p value < 0.05).*

Linear regression model

Multiple Model	Estimate (± SE)
(Intercept)	2.575 (± 0.441)***
CSD Domain	0.431 (± 0.352)
LSD Domain	1.141 (± 0.358)**
Repeat density	-2.616 (± 0.287)***
CG methylation	-0.410 (± 0.323)
$H3K4^{me3}$ chip peaks	0.034 (± 0.009)***
ATAC peaks	0.021 (± 0.007)**
Deletion density	0.050 (± 0.358)
Insertion density	-0.175 (± 0.234)
RI	0.175 (± 0.094)
R^2	0.48
Adjusted R^2	0.4779
Num. obs	2217

*** p < 0 ** p < 0.001

To compute regression, **collinearity** was evaluated using the **variance inflation factor** (VIF).

$$VIF_j = \frac{1}{1 - R_j^2}$$

LSD correlated with recombination frequency, contrary to CSD.
Repetitive DNA, histone methylation and accessible chromatin also showed a significant relationship with recombination (Table 1).

Table 1. *Regression model after VIF selection. Intercept of the model and slope for each predictor is indicated. Variables with VIF value higher than 5 were removed from the model (i.e. CHG methylation density and CDS density). SE: standard error.*

Figure 2. *Each distribution was paired-tested using Mann-Whitney U test (p value < 0.05).*

Relationship between recombination frequency and genomic and structural features

Recombination frequency **correlated positively** with coding sequence density (CDS), $H3K4^{me3}$ histone modification and accessible chromatin (ATAC peaks), and **correlated negatively** with repetitive DNA density (repeats) and DNA methylation (Fig. 2). No clear trend was observed for deletion and insertion of TEs.
Further subdivision of each distribution into structural domains showed that crossover occurrence was **consistently lower in CSD than in LSD**.

Conclusions

The regression model gave a preliminary estimate of the forces influencing recombination rate in grapevine, although correlations among some features prevented the determination of a main driving factor. Recombination was not significantly affected by structural variation due to TEs, suggesting that lack of sequence homology *per se* is not sufficient to explain crossover occurrence. Instead, such driving forces should be investigated more at the epigenetic regulation level.

Perspectives

By analysing higher number of progenies, we aim to reach a recombination frequency resolution at the single-gene level. This will allow to define recombination hotspots and to study genetic and epigenetic information in hotspot control by looking at promoters and other regulatory sites.

Supported by: "Novel variation in plant breeding and the plant pan-genomes - NOVABREED" Grant agreement n. 294780 - European Research Council