Exploring alternative splicing events in 10 different grapevine cultivars

Alessandro Cestaro

Computational Biology Group
alessandro.cestaro@fmach.it

October 1, 2013
Berries Transcriptome Atlas

One Tissue and One condition
- Transcriptome atlas at berries maturation, reduce the number of possible variables in a complex system
- from black to white, in skin and pulp, different sugar and anthocians accumulation influence vine quality

- RNA-seq: Illumina paired-ends reads
- 10 different *Vitis Vinifera* cultivars
- 20 million of reads each-one

*Pinot Noir, Teroldego, Alicante Bouchet, Sangiovese, Moscato Rosa, Lambrusco Salamino, Cabernet Franc Chardonnay, Inzolia, Poloskei Muskotaly*
Alternative Splicing in Plants

RNA-seq rapidly increased the fraction of gene that undergo AS

- in *Arabidopsis* the latest estimation reported 60% of multi-exonic genes
- in Rice similar RNA-seq analysis reported 40%

What bout the impact of AS isoforms?

- only few well documented examples
- functional role in the biotic and abiotic stress response
Whole Genome Alignment

Reference Genome
- Pinot Noir, cv. PN40024
- 19 Chr
- 500 Mb
- 2,068 scaffolds
- 14,657 contigs
- 28,268 gene predictions
We developed a new software

- other are mainly based on graph theory to explore all possible paths from all possible combinations of exons, full lengths isoform as final results
- If the production of an entire functional transcripts is not the principal role of AS, our analysis can suffer of unpredictable bias

Our approach

- Identify local events
- Low-abundance supported in multiple cultivars
findAS: pipeline

1. Genome Alignment
   - no matter the software, no matter the NGS source, just a BAM file

2. Primary clustering

3. Chimera search

"Local" Alternative Splicing Detection

4. AS detection

5. Evidence check
   - 3 cumulative reads from 3 different cDNA libraries
Splicing junctions (SJs)

- **107,330 of SJs in each cv (average)**
- In average 31% of the total SJs have been detected as novel junctions
- 92% of SJs are located inside the CDS

- Observed SJs confirmed the 95% of predicted genes
- The amount of new SJs is proportional to the amount of data
AS detection

- 40.4% of multiexonic genes
- Most common event is IR (37%); less common event ES (5%)
- Alt-5’, Alt’-3’ as a single category, reach the 57%
AS detection

3,735 AS events common to all cultivars
Majority of low-abundance events

An indication of the expression degree calculating the reads coverage of the alternative event divided by the coverage of the consensus form (Alternative Events Ratio).

- AS events have a mean AER value lower than 0.1
- $AER < 0.1$ : IR 72%, ES 89%, Alt-3’/Alt-5’ 74% and 72%
Alternative Junction relative position

Majority of Alt-3’ and Alt-5’ are very close to the annotated junction

- Alt-5’ $AER < 1$ prevalence for the positions not in frame (27.7% $P$-value $= 2.2e-16$)
Alternative Junction relative position

Majority of Alt-3’ and Alt-5’ are very close to the annotated junction

- Alt-5’ AER $< 1$ prevalence for the positions not in frame ($27.7\%$ P-value $= 2.2e-16$)
- Alt-3’ AER $\geq 1$ in frame position is prevalent ($43.7\%$ P-value $= 6.8e-9$)
AS and gene expression
AS and gene expression

![Graph 1: Freq. of AS events vs. gene exon number](image1.png)

- Equation: $a \log(x) + b$

![Graph 2: Mean number of predicted AS events vs. number of detected splicing sites / gene](image2.png)

- R squared: 0.76
- p-value: $6.4 \times 10^{-22}$
Conclusions

We have found evidence that at least 40% intron containing genes in berry undergo to AS

- All alternative spliced genes have always (90%) also the constitutive form.
- Majority of mRNA diversity observed derived by low-abundance events.
- Alt-3’ and Alt-5’ are really close to the canonical sites.
- AS frequency proportional to the gene exon content.
- AS extent is related to the expression level.
Conclusions

We have found evidence that at least 40% intron containing genes in berry undergo to AS

- All alternative spliced genes have always (90%) also the constitutive form.
- Majority of mRNA diversity observed derived by low-abundance events.
- Alt-3’ and Alt-5’ are really close to the canonical sites.
- AS frequency proportional to the gene exon content.
- AS extent is related to the expression level.

Stochastic noise affection in Alternative Splicing?
Stochastic noise in splicing machinery

E. Melamud et al, 2009

Number of Detected Splicing Reactions

Number of Alternative Splicing Reactions

$y = 0.15x^{0.5}$
Stochastic noise in splicing machinery

Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene

E. Melamud et al, 2009

M. Gonzalez-Porta et al., 2013
Acknowledgments

FEM
- Emilio Potenza
- Elisa Asquini
- Riccardo Velasco

Gent University & VIB
- Lieven Sterck
- Yves Van de Peer

University of Florence
- Milvia Racchi