Genome-wide association study of a diverse grapevine panel to uncover the genetic architecture of numerous traits of interest


AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France
Multiple changes and challenges

Reduce pesticides

Adapt to climate change

ref. $T^\circ$ in 1970
pred. $T^\circ$ in 2055
pred. $T^\circ$ in 2085

ARPEGE model

Major questions to biologists:
1. how to phenotype the eco-physiological processes of interest?
2. what are their genetic architectures?
3. how to incorporate them into breeding programs?
Multiple changes and challenges

Reduce pesticides

Adapt to climate change

Major questions to biologists:

1. how to phenotype the eco-physiological processes of interest?
2. what are their genetic architectures?
3. how to incorporate them into breeding programs?
Diversity panel of *Vitis vinifera* L. from Domaine de Vassal

Beside bi-parental populations $\Rightarrow$ 279 cultivars (weak structure)

Nicolas *et al.* (2016)
Field layout at Domaine du Chapitre

2009: overgraft on Marselan (control)

- 5 complete randomized blocks
- each genotype has 1 replicate per block

© 2009 AND, Tele Atlas, Google
Intense phenotyping effort

2010-2012

- Traits: mean berry weight; mean bunch weight, length and compactness; pruning weight and number of woody shoots; malate, tartrate, shikimate; $\delta^{13}$C
- Additional covariates: vigour, sanitary status
- No irrigation

2014-2015

- Traits: mean berry weight; $\delta^{13}$C; $\beta$-damascenone and pDMS; polyphenols (Pinasseau et al., 2017)
- Treatment: with or without irrigation
Intense phenotyping effort

2010-2012
  ▶ Traits: **mean berry weight**; mean bunch weight, length and compactness; pruning weight and number of woody shoots; malate, tartrate, shikimate; $\delta^{13}$C
  ▶ Additional covariates: vigour, sanitary status
  ▶ No irrigation

2014-2015
  ▶ Traits: mean berry weight; $\delta^{13}$C; $\beta$-damascenone and pDMS; polyphenols (Pinasseau et al., 2017)
  ▶ Treatment: with or without irrigation

⇒ Focus on **mean berry weight** (2010-2012)
Mean berry weight: exploratory analysis of phenotypes

Control genotype (Marselan) per block and year

--- outlier threshold

blocks and years


Flutre et al.

GWAS of grapevine
Mean berry weight: exploratory analysis of phenotypes

Panel per block and year

--- outlier threshold

blocks and years

A in 2010  B in 2010  C in 2010  D in 2010  E in 2010
A in 2012  B in 2012  C in 2012  D in 2012  E in 2012
Mean berry weight: exploratory analysis of phenotypes

Missing data in 2011
Dual genotyping

▶ **GrapeReSeq microarray** (Illumina): 12k SNPs after QC
▶ **GBS** with ApeKI enzyme (Keygene): 120k SNPs after QC
▶ **Combined**: **90k SNPs** with LD $< 0.9$ and MAF $> 0.01$
Dual genotyping

- **GrapeReSeq microarray** (Illumina): 12k SNPs after QC
- **GBS with ApeKI enzyme** (Keygene): 120k SNPs after QC
- **Combined**: 90k SNPs with LD < 0.9 and MAF > 0.01

11k SNPs

90k SNPs

⇒ **Densification** required to tag all/most causal polymorphisms
Kinship matrix from SNPs (additive genetic relationships)
Statistical analysis of phenotypic data

\[ y = X\beta + Zg + \epsilon \]  with  \( g \sim \mathcal{N}(0, \sigma_g^2 \text{Id}) \); \( \epsilon \sim \mathcal{N}(0, \sigma^2 \text{Id}) \)

- **\( y \):** phenotypic observations
- **\( \beta \):** effects of known factors, modeled as "fixed"
- **\( g \):** total genotypic values, modeled as "random"
- **\( \epsilon \):** errors
- **\( H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma^2 / \#\text{rep})} \):** broad-sense heritability (of means)
Statistical analysis of phenotypic data

\[ y = X\beta + Zg + \epsilon \text{ with } g \sim \mathcal{N}(0, \sigma_g^2 I_d); \epsilon \sim \mathcal{N}(0, \sigma^2 I_d) \]

- **y**: phenotypic observations
- **\beta**: effects of known factors, modeled as "fixed"
- **g**: total genotypic values, modeled as "random"
- **\epsilon**: errors
- \( H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma^2 / \#\text{rep})} \): broad-sense heritability (of means)

\[ y = X\beta + Za + \epsilon' \text{ with } a \sim \mathcal{N}(0, \sigma_a^2 A); \epsilon \sim \mathcal{N}(0, \sigma^2 I_d) \]

- **A**: kinship matrix of additive genetic relationships
- **a**: additive genotypic values (a.k.a. breeding values)
- \( h^2 = \frac{\sigma_a^2}{\sigma_g^2 + (\sigma^2 / \#\text{rep})} \): narrow-sense heritability (of means)
Estimation of heritabilities

$H^2$: higher, better $\to g$ well approximated by its BLUP

$h^2$: higher, better $\to \sigma^2_a$ large enough for selection purposes
Statistical analysis of genotypic values

SNP-by-SNP: *ad hoc*

\[
\text{BLUP}(g) = 1\mu + m_p \beta_p + u + \epsilon
\]

- \(\beta_p\): effect of the \(p^{th}\) SNP \(\rightarrow\) test if \(\beta_p = 0\)
- \(u\): polygenic effect with kinship matrix \(K \propto MM^T\)

Flutre et al. GWAS of grapevine
Statistical analysis of genotypic values

**SNP-by-SNP: ad hoc**

\[
\text{BLUP}(\mathbf{g}) = 1\mu + \mathbf{m}_p \beta_p + \mathbf{u} + \epsilon
\]

- \(\beta_p\): effect of the \(p^{th}\) SNP \(\rightarrow\) **test if** \(\beta_p = 0\)
- \(\mathbf{u}\): polygenic effect with kinship matrix \(K \propto \mathbf{M}\mathbf{M}^T\)

**Multi-SNP: explicit modelling of the genetic architecture**

\[
\text{BLUP}(\mathbf{g}) = 1\mu + \mathbf{M}\beta + \epsilon
\]
Statistical analysis of genotypic values

SNP-by-SNP: *ad hoc*

\[
\text{BLUP}(g) = 1\mu + m_p\beta_p + u + \epsilon
\]

- \(\beta_p\): effect of the \(p^{th}\) SNP \(\rightarrow\) test if \(\beta_p = 0\)
- \(u\): polygenic effect with kinship matrix \(K \propto MM^T\)

Multi-SNP: explicit modelling of the genetic architecture

\[
\text{BLUP}(g) = 1\mu + M\beta + \epsilon
\]

- fully polygenic: all \(\beta_p \neq 0\)
- major QTLs only: few \(\beta_p \neq 0\) and all others = 0
- hybrid: all \(\beta_p \neq 0\) and few \(\tilde{\beta}_p \neq 0\)
Estimation of hybrid genetic architectures

**PVE**: proportion of variance of total genotypic values explained by the polygenic component and the major QTL effects

- higher → better to predict genotyping values

**PGE**: proportion of PVE explained *only* by major QTL effects

- higher → better to identify candidate genes
Estimation of hybrid genetic architectures

**PVE**: proportion of variance of total genotypic values explained by the polygenic component and the major QTL effects
- higher → better to predict genotyping values

**PGE**: proportion of PVE explained only by major QTL effects
- higher → better to identify candidate genes

<table>
<thead>
<tr>
<th>trait</th>
<th>#SNPs</th>
<th>med(#QTLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mbw 11k</td>
<td>11k</td>
<td>31 [0,169]</td>
</tr>
<tr>
<td>mbw 90k</td>
<td>90k</td>
<td>14 [2,115]</td>
</tr>
</tbody>
</table>

- importance of genotyping densification
- large amount of genetic variance from polygenic component
Mean berry weight: SNP-by-SNP versus multi-SNP

SNP-by-SNP with 11k SNPs

⇒ genotyping not dense enough
Mean berry weight: SNP-by-SNP versus multi-SNP

SNP-by-SNP with 90k SNPs

⇒ dense enough to find two significant SNPs
Mean berry weight: SNP-by-SNP versus multi-SNP

Multi-SNP (major QTLs only) with 90k SNPs

⇒ more power to find six SNPs tagging putative QTLs
Mean berry weight: SNP-by-SNP versus multi-SNP

Focus on the selected SNPs

\[ \hat{\text{PVE}} = 0.668 \ [0.613, 0.735] \]

- need to define QTLs around selected SNPs
Mean berry weight: selected SNPs

SNP #1 at \( \approx 6.3 \) Mb on chr17 (overlap known QTLs)

\[
\Pr(\beta_p \neq 0) = 1 ; \text{PVE}_p = 0.094 ; \hat{\beta}_p = -0.213 ; \text{CI}_{95\%} = [-0.263, -0.163] \\
\]

location: coding of Vitvi17g00537, (-)-isopiperitenol/(-)-carveol dehydrogenase, mitochondrial
Mean berry weight: selected SNPs

SNP #2 at \( \approx 29.9 \) Mb on chr14

\[ \Pr(\beta_p \neq 0) = 0.999 ; \hat{PVE}_p = 0.074 ; \hat{\beta}_p = -0.159 ; \text{CI}_{95\%} = [-0.202, -0.117] \]

Location: promoter of Vitvi14g02008, uncharacterized
Prospects with the panel

Phenotyping:
- improved phenotyping of berry physiology (poster 49); tolerance to pathogens (poster 57)
- phenotyping in multiple sites and greenhouses to study GxE

Genotyping:
- capture-based sequencing of GBS-defined SNPs
- search for traces of selection

Modeling:
- genomic prediction to speed-up selection (poster 82)
- multi-pop/-trait statistical analysis (ongoing work)
Take-home message

With **dense** genotyping and **multi-SNP** models, the **diversity panel** of *V. vinifera* L. from INRA Montpellier allows estimating the **genetic architecture** of numerous traits of interest, to help design efficient **breeding** strategies.

- diversity panel: virus-free and available
- data and reproducible analyzes: available upon publication
- contact: Agnès Doligez (agnes.doligez@inra.fr)
Acknowledgments

▶ DAAV team from the UMR AGAP
▶ Vassal-Montpellier grapevine biological resources center
▶ SouthGreen bioinformatics platform (notably B. Pitollat)
▶ AGAP genotyping platform (notably P. Mournet)
▶ GenoToul sequencing center
▶ all contributors to free/open softwares used in this work
▶ funding bodies: INRA, Montpellier SupAgro, IFV, ANR, CASDAR, EU