The Grapevine Genome and Beyond

A-F Adam-Blondon
Unit of research on Plant Genomics (URGV)
Unit of research on Genomics-informatics (URGI)

3rd Int Conf on Omics and Biotech in Fruit crops, Nanjing, 2012 October 28th
Grapevine is a crop with a high economic value

- 27.9 billion $ value for worldwide wine exports (8.4 billion $ for France) in 2010
- 4,601 million $ for worldwide grape and grape juices exports in 2006

http://faostat.fao.org/
Grapevine is a crop with a high economic value

- Grape producers in all continents

- Grapevine is adapted to poor soils and low water availability
Viticulture in Europe faces major challenges

1- Decrease the use of pesticides

• The viticulture has to reduce its use in pesticides because of:
  • Citizen awareness of fungicide effects on health and environment
  • EU policies aiming at reducing the sources of water pollution
  • State policies aiming at reducing the environment and health risks

• In France (2006 studies: « Ecophyto R&D », « Pesticides »):
  • France was the third worldwide consumer of pesticides even though a decrease has been observed since 2001.
  • 14% the pesticide value (20% of active compounds) was used in vineyards although they represent 3% of the cultivated surface
  • 13 to 21 treatments/years, from which 10 to 18 were fungicide applications (control of mildews and botrytis)
Viticulture in Europe faces major challenges

1- Decrease the use of pesticides

- High variability of practices for disease control among the vineyards:
  - possibility to spread the best one
  - Possibility to develop better methods to survey and control epidemics

- Improvement of the machines used for spraying: only 40-60% of the sprayed products end on leaves

- Development of highly qualitative resistant varieties:
  - Construction of durable resistance while maintaining high quality

- Combine everything to increase the security of the producers and the durability of the resistances
Viticulture in Europe faces major challenges

1- Decrease the use of pesticides

Construction of durable resistances

- Phenotyping for disease resistance
- Genomic tools for the study of the evolution of the pathogen populations
- Marker Assisted Breeding at a genomic scale
- Diverse phenotyping
- Identification of sources of resistance and characterisation of their effect on pathogens
- Pyramiding into a *Vitis vinifera* background in order to construct a durable resistance
- Breeding for quality, adaptation to environment
- The combinations of quantitative and qualitative resistances are more difficult to overcome by the pathogen (*Palloix et al* New Phytol 2009)
- Run1, Ren1, Rpv1 to 8, Pdr1a, Pdr1b, ...
- Assessment of durability

*Breuer&Milgroom 2010 Peressotti et al 2010*
Viticulture in Europe faces major challenges

1- Decrease the use of pesticides

- Construction of durable resistances
  - Phenotyping for disease resistance
  - Genomic tools for the study of the evolution of the pathogen populations
  - Marker Assisted Breeding at a genomic scale
  - Diverse phenotyping
  - Identification of sources of resistance and characterisation of their effect on pathogens
  - Pyramiding into a *Vitis vinifera* background in order to construct a durable resistance
  - Breeding for quality, adaptation to environment
  - Assessment of durability

- Access and exchanges of GR limited and complex (biology, sanitary issues, international treaty)

- How can we manage the durability of the resistance at the level of a landscape/region?
Viticulture in Europe faces major challenges
2- Adaptation to climate change

Hypothesis:
Constant rate of increase over 50 ans

From S. Delrot
Viticulture in Europe faces major challenges
2- Adaptation to climate change

Phenological changes observed in the Cognac area (Ugni Blanc)

- Shortening of the period between budbreak and harvest
- Earlier phenological stages
- Ripening phase switched to warmer summer periods

(G.Snakkers, BNIC 2007)

Ethanol content in wines from the Alsace area (Riesling)

- Over 2 degrees gained in 35 years
- However there are detrimental effects of the heat on the aromatic quality of wine in some areas.

(SIVA statistics)
Viticulture in Europe faces major challenges
2- Adaptation to climate change

- Combination of environmental parameters affected
  - CO₂ level
  - Température
  - Hygrometry
  - Soil water content

- Plant processes affected (that may interact with each other)
  - Phenology
  - Yield potential
  - Fruit composition
  - Response to pathogens

- Major physiological targets that may be affected
  - Photosynthesis/Respiration
  - Stomatal regulation
  - Hydraulic conductance
  - Growth of organs and carbon allocation
  - Development
  - Ion uptake and assimilation
  - Primary and secondary metabolism

From S. Delrot
Viticulture in Europe faces major challenges

2- Adaptation to climate change

Genetic variability for adaptation to environment

Phenotype = G + E + G×E
P = Gv + Gr + E + GvGr + GvE + GrE + GvGrE

(Gv = genotype of variety; Gr = genotype of rootstock)

(from Jones, 2006)
Viticulture in Europe faces major challenges

2- Adaptation to climate change

- How to predict a phenotype from the informations available on the genotype and its interactions with the environment?
  > How will the existing varieties respond, and what are the new phenotypes we are looking for (ideotypes)?

Modeling at different scales

Tardieu&Tuberosa 2010
Marguerit et al 2012

Cell Organ Plant Vineyard Region/landscape
Viticulture in Europe faces major challenges

2- Adaptation to climate change

Phenotyping at different scales

Genomics to help understand the underlying processes

From S. Delrot
Reference genome sequence of *V. vinifera* as a basic tool

- Access to the about 30,000 genes and their order along the genome in the *Vitis* genus for genetic and functional analysis
- Possibility to quickly develop markers in any genome region in the *Vitis* genus

8X version of the genome published in 2007 (Jaillon et al 2007)
## Current status: 12X.0 assembly

<table>
<thead>
<tr>
<th>Library</th>
<th>Average insert length</th>
<th>Number of reads</th>
<th>Number of mate pairs</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>100k</td>
<td>120091</td>
<td>57,606</td>
<td>0.12x</td>
</tr>
<tr>
<td>Fosmids</td>
<td>40k</td>
<td>251425</td>
<td>117,869</td>
<td>0.34x</td>
</tr>
<tr>
<td>Plasmids</td>
<td>10k</td>
<td>2848550</td>
<td>1,387,612</td>
<td>3.77x</td>
</tr>
<tr>
<td>Plasmids</td>
<td>3k</td>
<td>5518286</td>
<td>2,705,911</td>
<td>7.68x</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>8738352</td>
<td>4,268,998</td>
<td>11.91x</td>
</tr>
</tbody>
</table>

### Sanger shotgun sequencing approach

```
contig
```

```
super-contig
```

12X.0 version (consensus map from 2 largers pop; 514 unique SSR markers):
• 91% of the draft sequence is anchored on the genetic map
• 87% is anchored and oriented


http://urgi.versailles.inra.fr/
12X.0 : characteristics of the remaining « random » sequence

Anchored to a chromosome but not oriented : 38 sgtg; 16.8 Mb

- Anchored and oriented sequence : 2.33% of ‘N’
- 4.33% of ‘N’

Not anchored to a chromosome : 44 of them >200kb; 15.6 Mb

- 10.22% of ‘N’
- 2064 gene models (7.8%)
Improvement of the genetic map

Integrated SxG and ChxBi (Cipriani et al 2011)

A. Canaguier
Anchoring the scaffolds using the new genetic map

V0: 426 Mb anchored (85% genome) -> V1: 436.8 Mb anchored (87% genome)

V0: 355 Mb oriented (71% genome) -> V1: 415 Mb oriented (83% genome)
Anchoring the scaffolds using the new genetic map
Current Annotation from the CRIBI: V1 annotation

Input:
- 339,008 Vitis ESTs (NCBI)
- flcDNAs (99,828 reads from 5 libraries; URGV + Genoscope)
- Deep EST sequencing using Illumina-Solexa (175M reads from 4 libraries; IGA) and Roche-454
- V0 automatic annotation using the Gaze software (Genoscope)

• 12X.0 annotation V1 : 29 971 gene models => used for the whole genome Nimbelgen transcriptome array (Univ. Verona)
• Repeat annotation (IGA + CRIBI)

Proposition of a stable automatic gene numbering based on the Arabidopsis system (AtXgZZZZZ):
VvXX  sYYYY  gZZZZ
n°chr  n°super-contig  n°gene on super-contig
Several important gene families have been expertised, both at the level of the gene model and of the function:

- 152 genes/pseudogenes of the terpene synthase gene family (Martin et al 2010; [http://urgi.versailles.inra.fr/gb2/gbrowse/vitis_12x_pub](http://urgi.versailles.inra.fr/gb2/gbrowse/vitis_12x_pub))
- 829 genes/pseudogenes of the NSB-LRR gene family (unpublished yet)
- 19 genes putative TF with a TIFY domain (Zhang et al 2010)
- 17 PR10 related genes/pseudogenes (Lebel et al 2010)
- 17 CDPK genes (see the presentation of Pr Cheng)
- Etc...

Toward a V2 annotation in the frame of a new EU COST action in collaboration with the tomato genomic consortium.
Improvement of the V1 annotation

http://urgi.versailles.inra.fr/index.php/urgi/Species/Vitis/Resources/Genome-sequence-and-annotations

Apollo (login required) -> Chado
Write

Chado -> Export Edited genes (GFF3)

Chado -> Finished genes

New complete release

Chado -> In progress genes

Chado -> Problematic genes

Apollo WebStart

Apollo (login required) -> Gbrowse update
Read

Gbrowse update -> Bio:Seq: Feature store
Read

Bio:Seq: Feature store

Never used! Too complex?
Improvement of the V1 annotation

- Automatic annotation by IGA (Udine, Italy)
- Curation of the Class II transposons (Benjak et al 2008, 2009)

**REPET V2** (http://urgi.versailles.inra.fr)

Novel TE automatic annotation using REPET V2 (Flutre et al 2011)

First round: 65% genome = TE
- 41.4% Class I
- 8.7% Class II
- 33.8% unclassified repeats according to Wicker et al 2007

=> 7822 consensus

Second round:
- 6738 consensus with at least one complete copy in the genome
- => 67.8% genome = TE

Manual curation on going
Where can you access all these informations?

http://urgi.versailles.inra.fr

Access to GnipIS portal

Access to Vitis resources
Markers

27 Vitis maps
Integrated_2 (Cipriani et al 2011)

Vitis web pages

12X.0 GBrowser

• V0 and V1 annotations
• Curated TPS
• BES alignment (PN4014, CS, soon Muscadine)

Polymorphism data

• 27 Vitis maps
• Integrated_2 (Cipriani et al 2011)

Transcriptomics data (GRASP project)

• SNPs: Myles et al 2011
• Houel et al 2010
• Houel et al unpublished

SiREGAL
Genetic resources db

Passeport data of the grapevine National collection (> 7000 accessions)

GnpSNP Polymorphism db

> 1.6 million variants
Molecular basis of trait variation and development of MAS at a genomic scale
Illumina re-sequencing of 47 grapevine genotypes

In the frame of EU KBBE GrapeReSeq project and of a french ANR project Muscares

Leaf harvest and DNA prep

DNA quantification
Library prep
Illumina GAII sequencing

Triming for quality

Alignment on the reference Genome sequence
SNP detection

20 diverse Vitis vinifera
4 Vitis sylvestris
3 V. cinerea
3 V. berlandieri
3 V. aestivalis
3 V. labrusca
6 M. rotundifolia

French National plateform for Genotyping (CNG), Evry, France

Results from the re-sequencing of 25 V. vinifera varieties at IGA (Udine, Italy)

Illumina 20K genotyping Chip (http://urgi.versailles.inra.fr => Vitis pages)
Pipeline for SNP detection: MAPHiTS
Mapping Analysis Pipeline for High-Throughput Sequencing

- Short Reads File 1
- Short Reads File 2
- Reference File

Header Fasta Filter

VarScan Filters:
- allele frequency
- pValue
- base quality...

SAMtools
- Sam to Bam
- Bam to Pileup
- Bam to Tablet
- Bam to GenomeView

GnpSNP db
- Export to URGIdatabase

VarScan

Export to URGIdatabase

Calculations parallelized
MAPHiTS workflow in Galaxy
(a web-based platform for genomic research: http://usegalaxy.org)
### Sequencing results for the 47 V. vinifera varieties

<table>
<thead>
<tr>
<th>Read Length</th>
<th>Average</th>
<th>Standard dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read length (after trim)</td>
<td>94</td>
<td>6,38913</td>
<td>72</td>
<td>100</td>
</tr>
<tr>
<td>Bases after trimming</td>
<td>9 245 106 087</td>
<td>6 109 483 264</td>
<td>3 717 274 733</td>
<td>33 989 041 742</td>
</tr>
<tr>
<td>Genome coverage after trimming</td>
<td>19</td>
<td>12.6</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>Genome coverage with uniq reads</td>
<td>13</td>
<td>8.1</td>
<td>4</td>
<td>47</td>
</tr>
<tr>
<td>% of the genome covered</td>
<td>81%</td>
<td>5%</td>
<td>67%</td>
<td>93%</td>
</tr>
<tr>
<td>SNP detected</td>
<td>4 285 569</td>
<td>1 798 766</td>
<td>734 738</td>
<td>8 661 975</td>
</tr>
<tr>
<td>SNP after filters</td>
<td>926 880</td>
<td>336 957</td>
<td>162 806</td>
<td>1 545 045</td>
</tr>
</tbody>
</table>

• GA II and/or High Seq sequencing
• Filtering for position in known structural variants, in repeats, with too few or too many reads
The GrapeReSeq_20K Illumina array:

• 15,022 SNP from *Vitis* vinifera
• 4,978 SNP from *Vitis* species

SNP chosen on the following criteria:

• Illumina score $\geq 0.9$
• SNP for Infinium type II array
• MAF criteria:
  • 90% SNP with MAF $> 0.1$ (85,463 SNPs)
  • 10% SNP with $0.05 < \text{MAF} < 0.1$ (27,631 SNPs)
• Heterozygosity (*Vitis* species SNP)
• Genome distribution
Aknowledgements

URGI – INRA de Versailles
   F. Alfama
   J. Amselem
   S. Arnoux
   M. Bras
   N. Choisne
   S. Durand
   T. Flutre
   O. Inizan
   V. Jamilloux
   N. Lapalu
   I. Luyten
   N. Mohellibi
   S. Reboux
   D. Steinbach
   D. Valdenaire
   D. Verdelet
   H. Quesneville

AGAP – INRA Montpellier
   P. This
   R. Bacilieri
   V. Laucou

SVQV – INRA Colmar
   P. Mestre

URGV – INRA Evry
   A-F Adam-Blondon
   A. Canaguier
   C. Houel
   I. LeClainche
   P. Faivre-Rampant
   C. Guichard
   S. Aubourg

ISVV – Bordeaux
   S. Delrot
Aknowledgements

IGA/ University of Udine, Italy
Michele Morgante
Gabriele Di Gaspero
Simone Scalabrin

University of Verona, Italy
Mario Pezzotti
Massimo Delledonne

University of Padova, Italy
Giorgio Vallé
Nicola Vitulo

JKI, Germany
L. Hausmann
R. Töpfer

University of Bielefeld, Germany
B. Weissaahr

CIVV, Spain
J-M Martinez-Zapater
Thank you!
Mapping with 99% of identity

Distribution of the mapped (blue) and unmapped (red) paired reads

Muscadinia rotundifolia
Anchoring the scaffolds using the new genetic map

Nearly ready to provide a new chromosome sequence

Wang et al 2012
Improvement of the genetic map

• Design of **200 SNP** markers from the sequence of 37 non oriented scaffolds and 40 large non anchored scaffolds
• Recovery of SNP genotyping data from Vezzulli et al 2008
• Recovery of SSR genotyping data from Doligez et al 2006
• Recovery of SSR genotyping data from Cipriani et al 2011

• **1022** SSR primer pairs/SNP loci used on two populations:
  • **535** markers in the Syrah x Grenache population (193 indiv)
  • **730** markers in the Chardonnay x Bianca population (358 indiv)

A. Canaguier
System Biology of adaptation to biotic and abiotic stresses

- Need to combine modeling at different scales and subjects and taking into account genetic variation

- Need connected databases with different layers of information

- System biology of two organisms in interaction…
  => Need genomic tools for the pathogens (genome sequence)

- …or may be more (wood diseases, new decay syndromes…)
  => Metagenomics approaches