



HAL
open science

Grapevine genome and beyond

Anne-Francoise Adam-Blondon

► **To cite this version:**

Anne-Francoise Adam-Blondon. Grapevine genome and beyond. 3. International Conference on Omics and Biotechnology in Fruit crops, Oct 2012, Nanjing, China. hal-02803277

HAL Id: hal-02803277

<https://hal.inrae.fr/hal-02803277>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Grapevine Genome and Beyond



B. Londinsky



A-F Adam-Blondon

Unit of research on Plant Genomics (URGV)

Unit of research on Genomics-informatics (URGI)

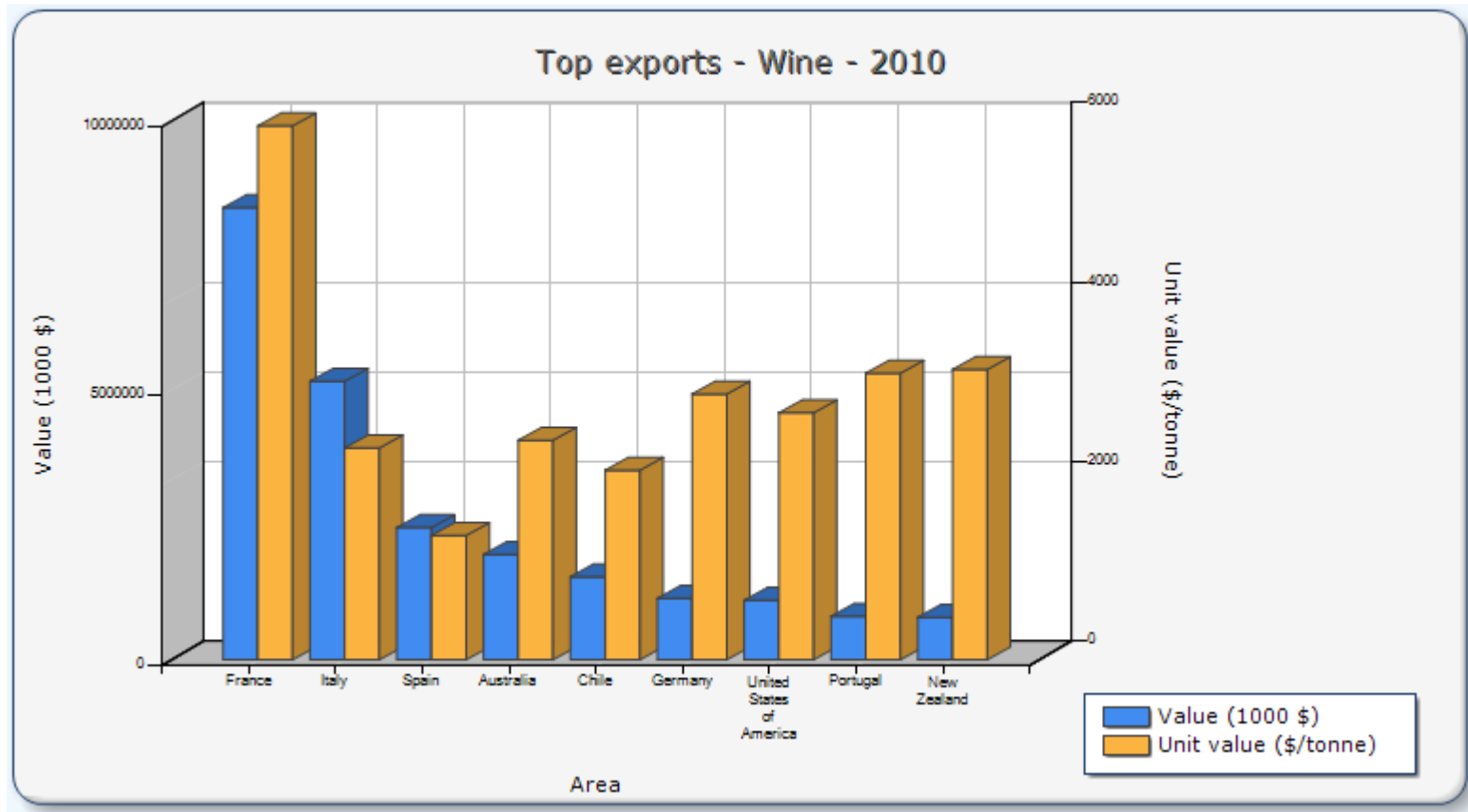


3rd Int Conf on Omics and Biotech in Fruit crops, Nanjing, 2012 October 28th

Grapevine is a crop with a high economic value

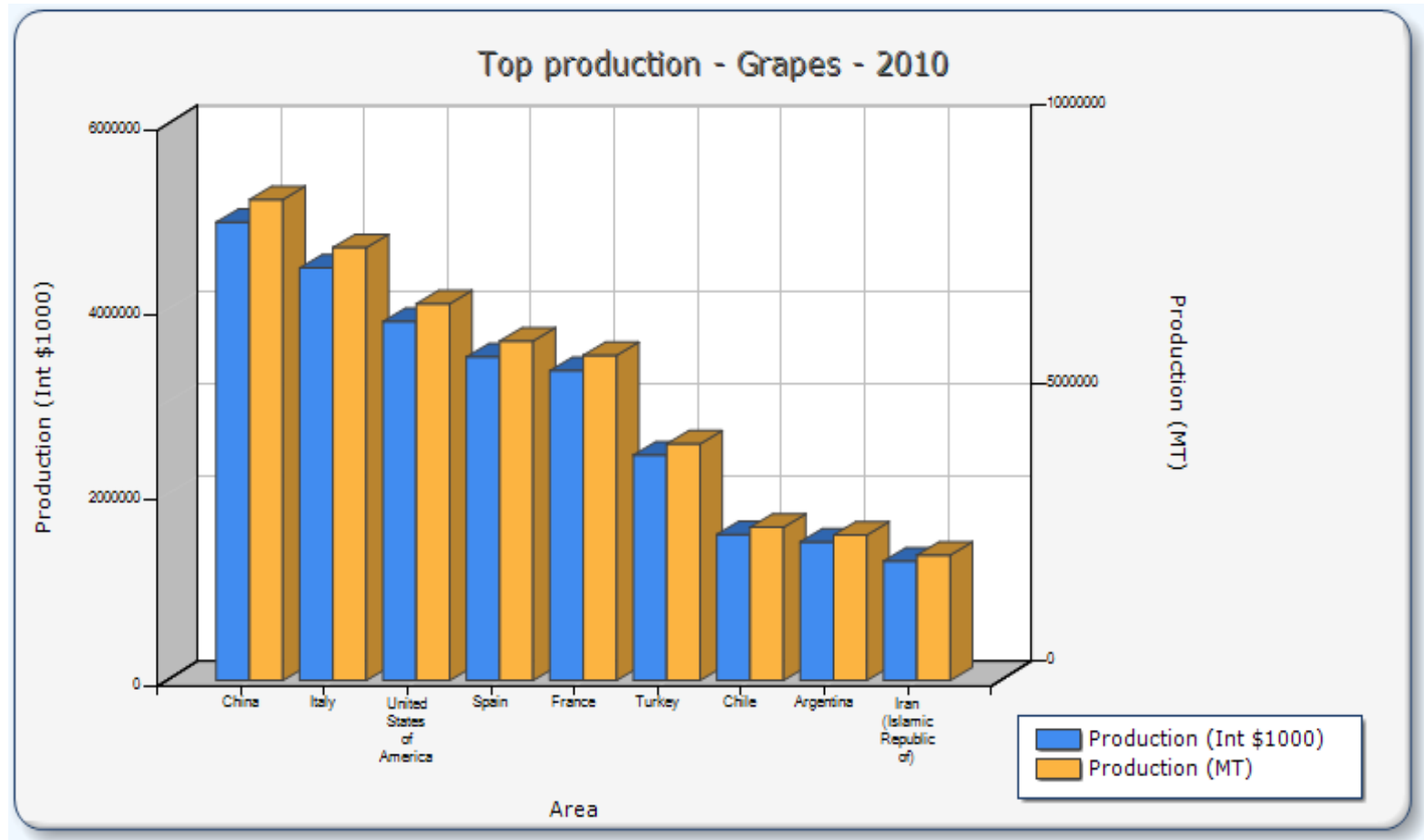
- 27.9 billion \$ value for worldwide wine exports (8.4 billion \$ for France) in 2010
- 4,601 million \$ for worldwide grape and grape juices exports in 2006

<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

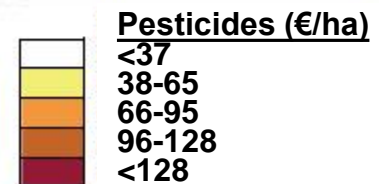
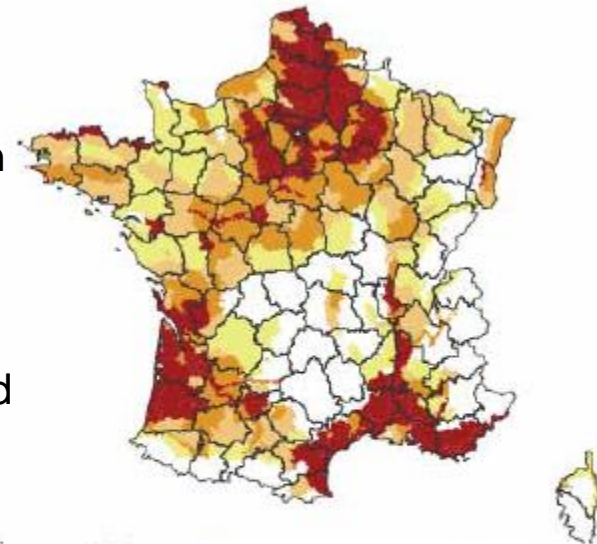
Viticulture in Europe faces major challenges

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

Run1, Ren1, Rpv1 to 8, Pdr1a, Pdr1b, ...



Assessment of durability

Brewer&Milgroom 2010
Peressotti et al 2010

Pyramiding into a *Vitis vinifera* background in order to construct a durable resistance

The combinations of quantitative and qualitative resistances are more difficult to overcome by the pathogen (**Palloix et al New Phytol 2009**)



Breeding for quality, adaptation to environment

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



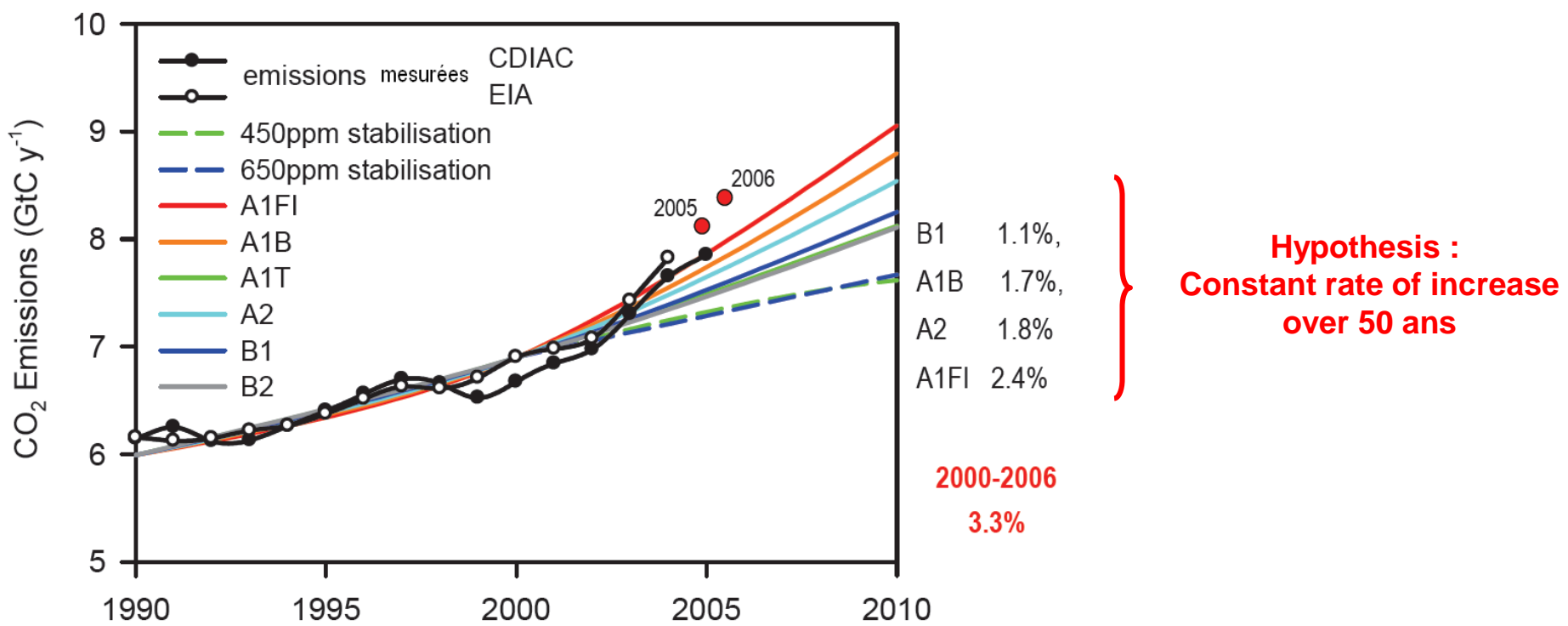
Assessment of durability

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

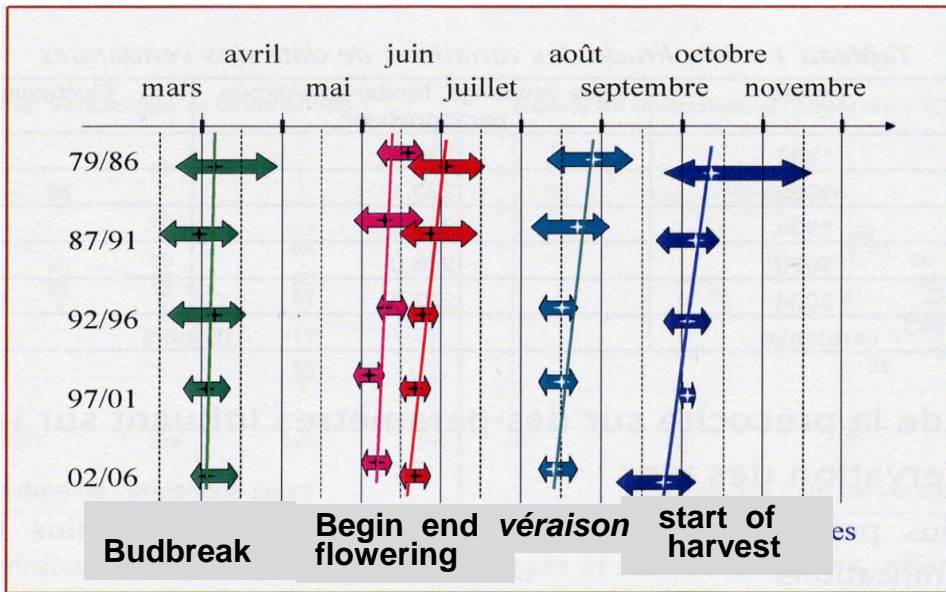
2- Adaptation to climate change



From S. Delrot

Viticulture in Europe faces major challenges

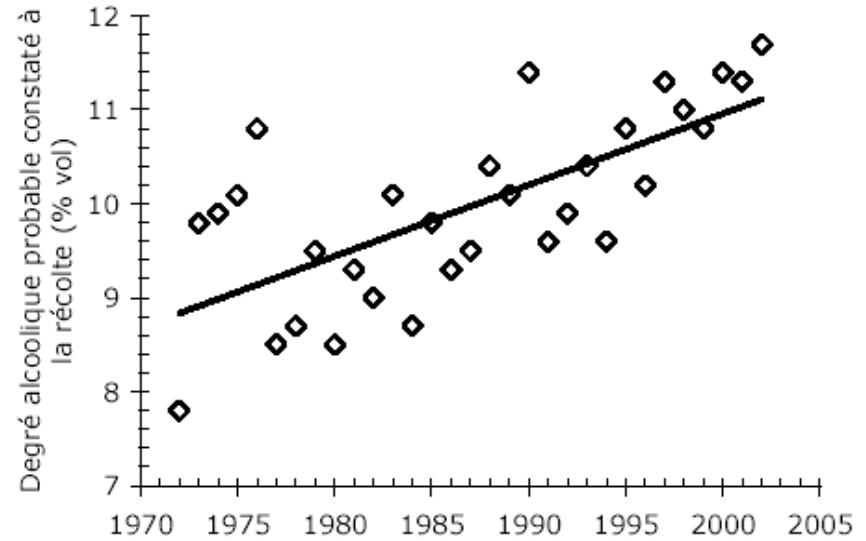
2- Adaptation to climate change



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.Snackers, BNIC 2007)



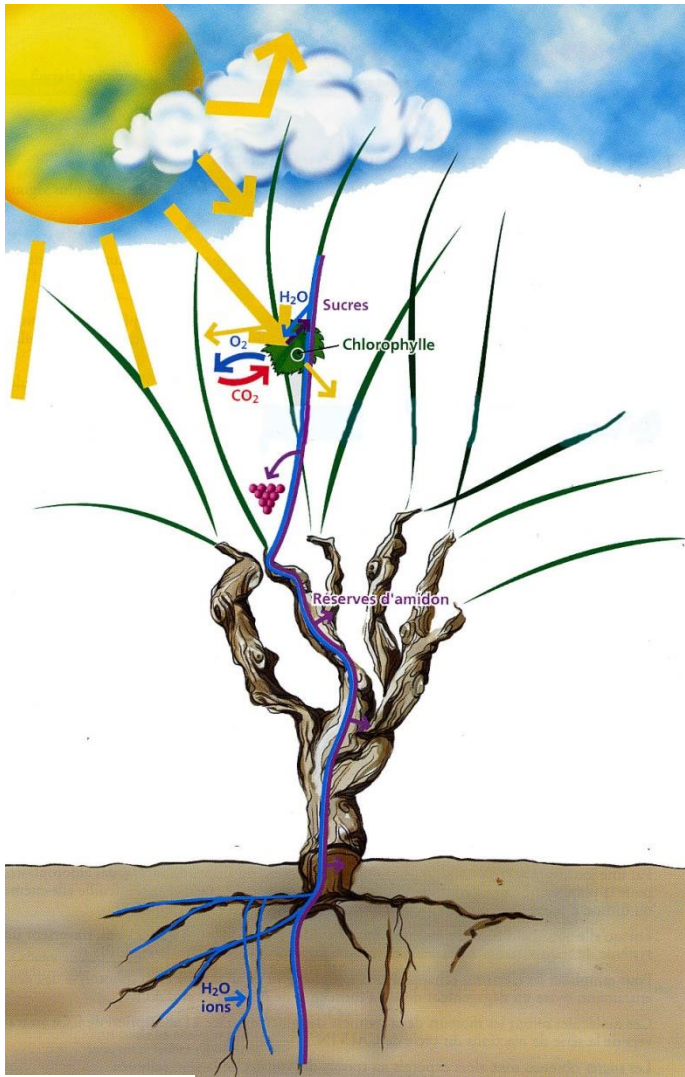
Ethanol content in wines from the Alsace area (Riesling)

- 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

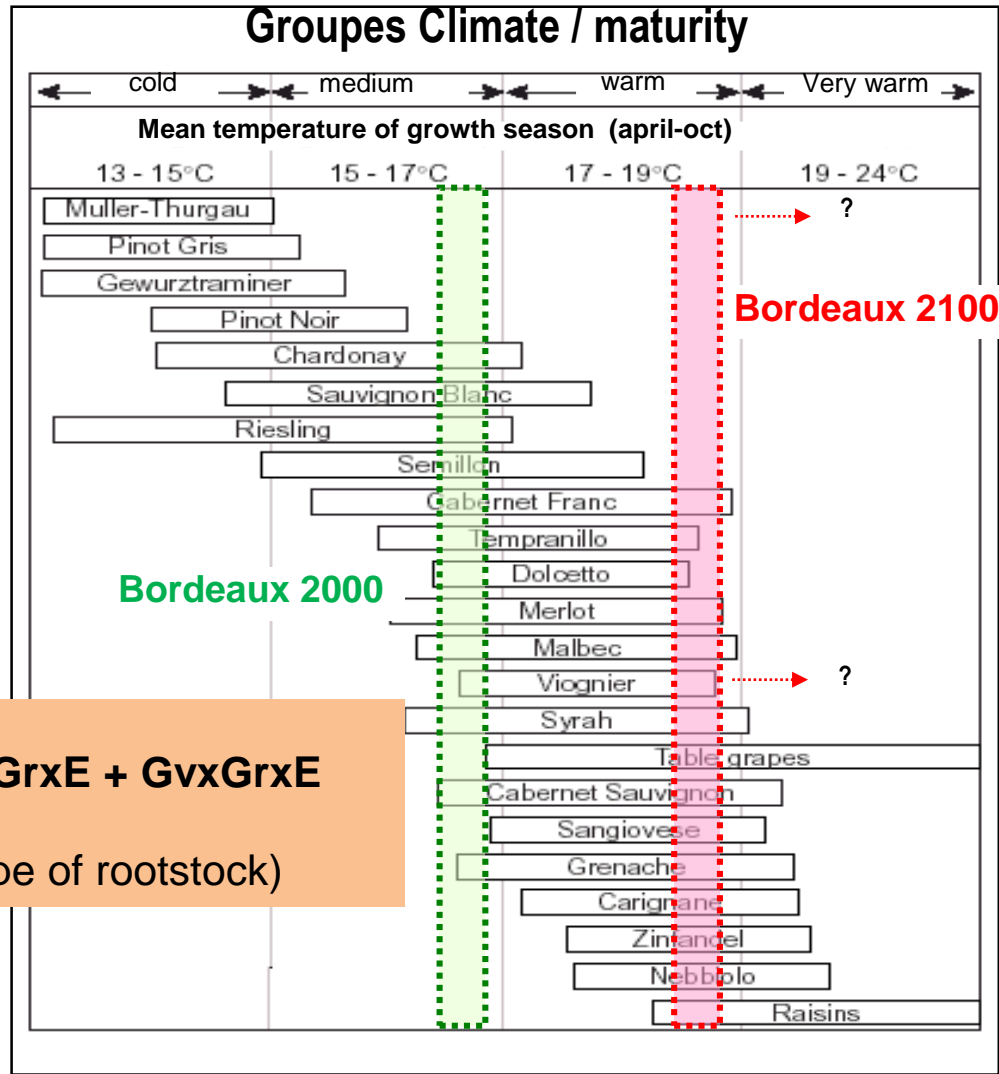


- **Combination of environmental parameters affected**
 - CO_2 level
 - Température
 - Hygrometry
 - Soil water content
- **Plant processes affected (that may interact with each other)**
 - Phenology
 - Yield potential
 - Fruit composition
 - Response to pathogens
- **Major physiological targets that may be affected**
 - Photosynthesis/Respiration
 - Stomatal regulation
 - Hydraulic conductance
 - Growth of organs and carbon allocation
 - Development
 - Ion uptake and assimilation
 - Primary and secondary metabolism

Viticulture in Europe faces major challenges

2- Adaptation to climate change

Genetic variability for adaptation to environment



$\text{Phenotype} = G + E + G \times E$
 $P = G_v + G_r + E + G_v \times G_r + G_v \times E + G_r \times E + G_v \times G_r \times E$
 (Gv = genotype of variety; Gr = genotype of rootstock)

Viticulture in Europe faces major challenges

2- Adaptation to climate change

From S. Delrot



- 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

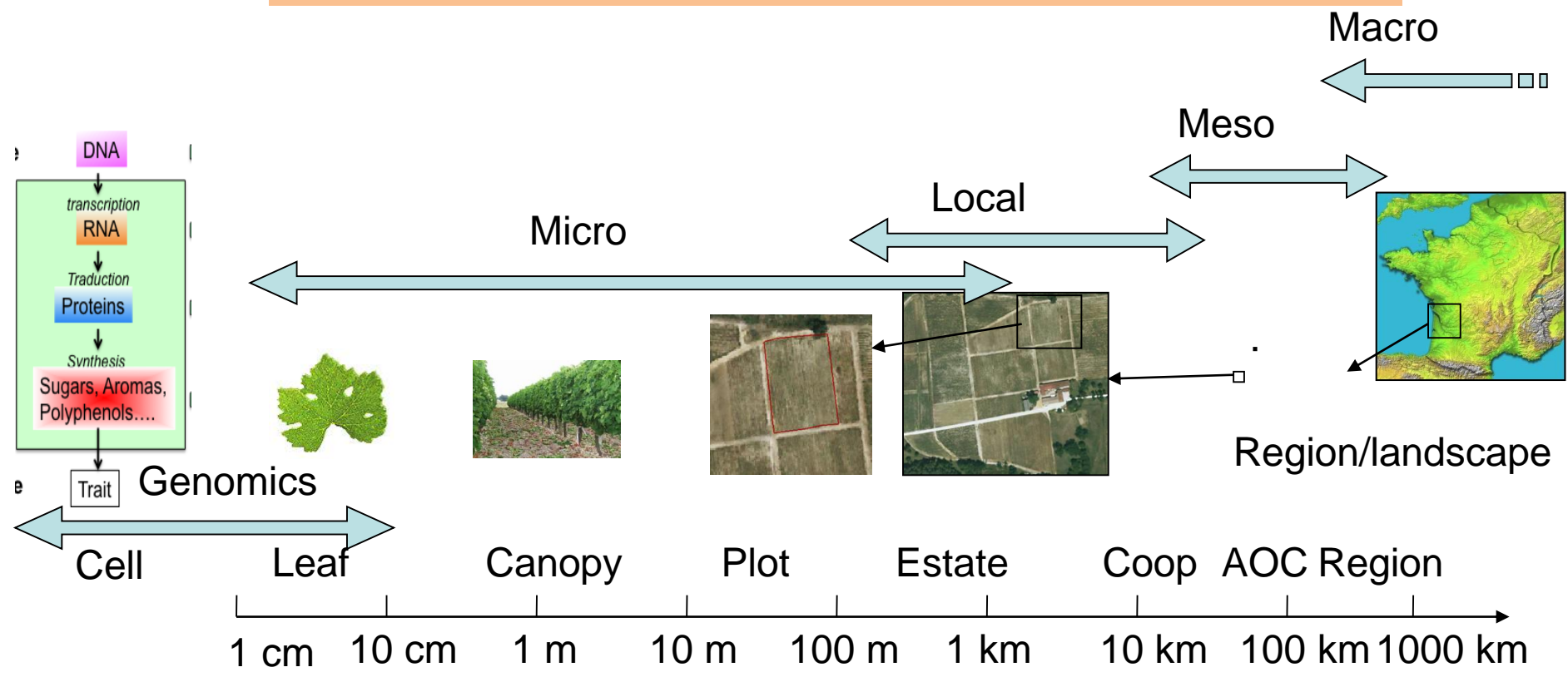


Region/landscape

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

Sequencing the grapevine genome

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

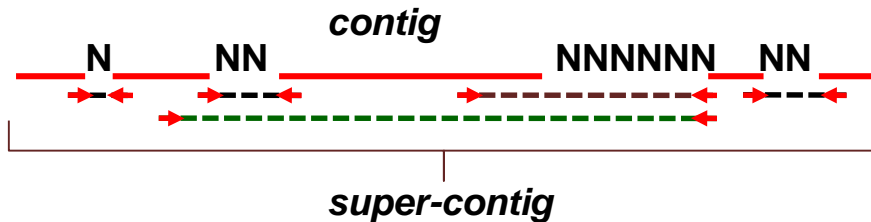
- Access to the about 30,000 genes and their order along the genome in the *Vitis* genus for genetic and functional analysis
- Possibility to quickly develop markers in any genome region in the *Vitis* genus

8X version of the genome published in 2007 (Jaillon et al 2007)

Current status: 12X.0 assembly

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

Sanger shotgun sequencing approach



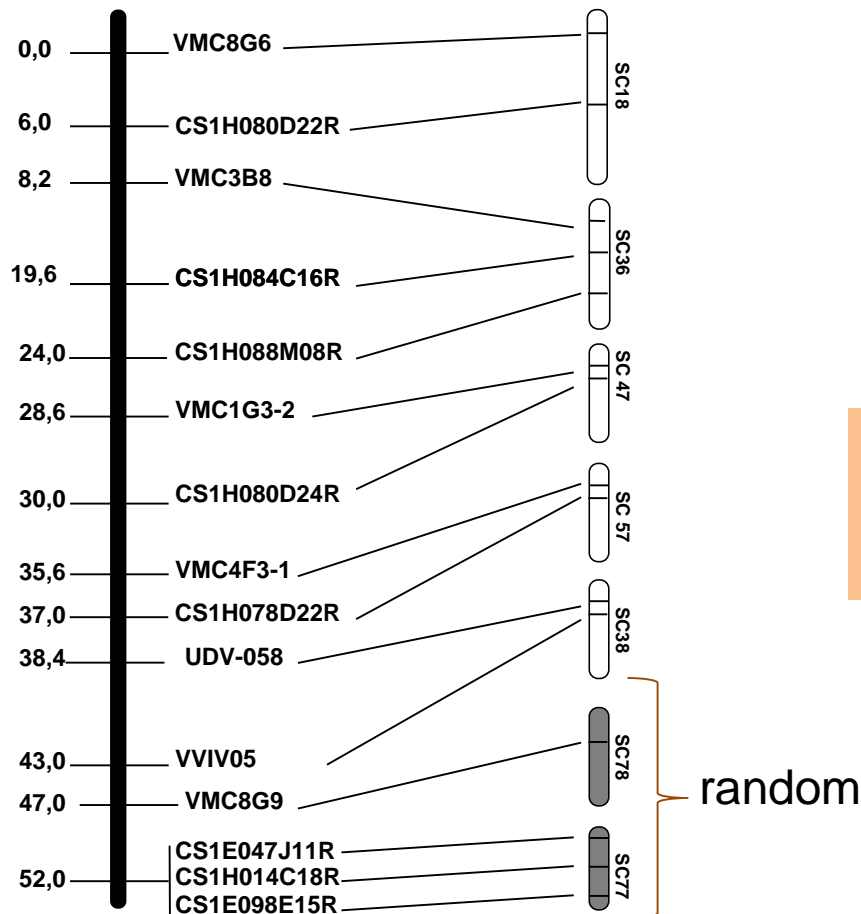
Adam-Blondon et al (2011) Genome Sequence Initiatives. *in: A-F Adam-Blondon, JM Martinez-Zapater, Chittaranjan Kole (eds) Genetics, Genomics and Breeding of Grapes.* Science Publishers and CRC Press. pp 211-234

	8X	12X
Contigs Nb	19,578	17,196
Longest Ctg (kb)	557	774
Ctg N50 (kb)	65	100
Total size	467	481
Super-contig Nb	3,515	2,888
Longest sctg (kb)	12,675	13,101
Sctg N50 (kb)	2,065	3,300
Total size	487	499

Sequence built of the 19 chromosomes

Genetic map :
chromosome 12

Super-contigs of
sequence



**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

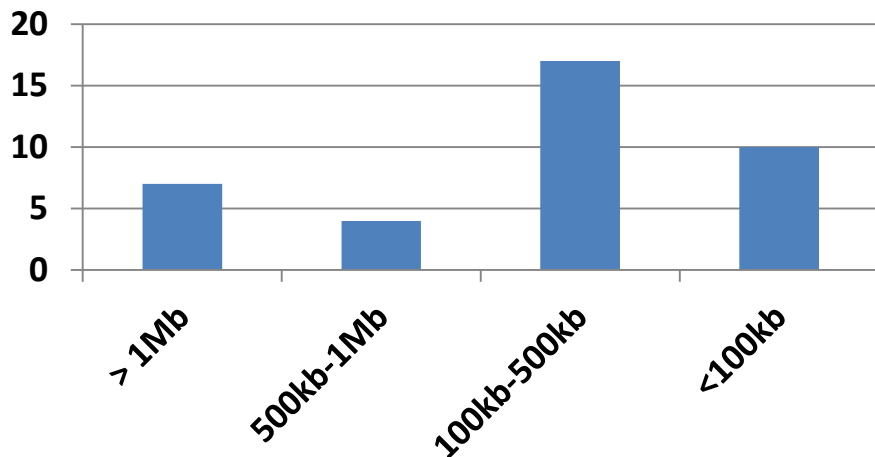
Cipriani et al (2011) Molecular Linkage Maps: Strategies, Resources and Achievements. *in*: A-F Adam-Blondon, JM Martinez-Zapater, Chittaranjan Kole (eds) *Genetics, Genomics and Breeding of Grapes*. Science Publishers and CRC Press. pp 111-136

<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

n° super-contigs

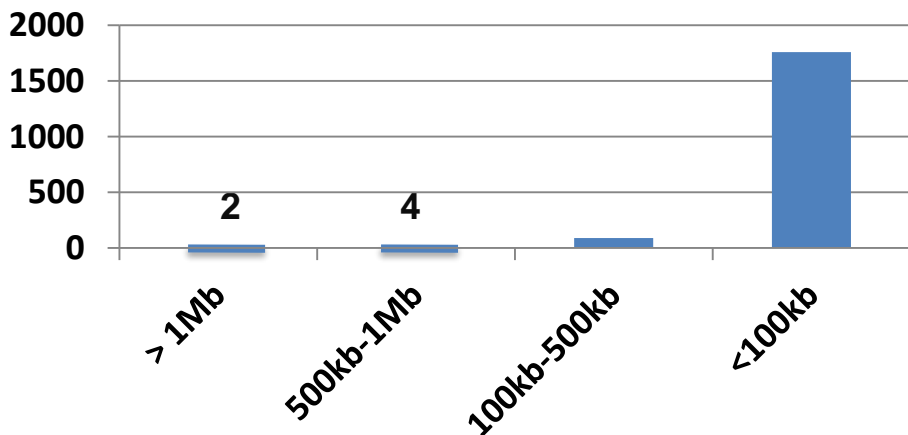


4.33% of 'N'

Anchored and oriented sequence : 2.33% of 'N'

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

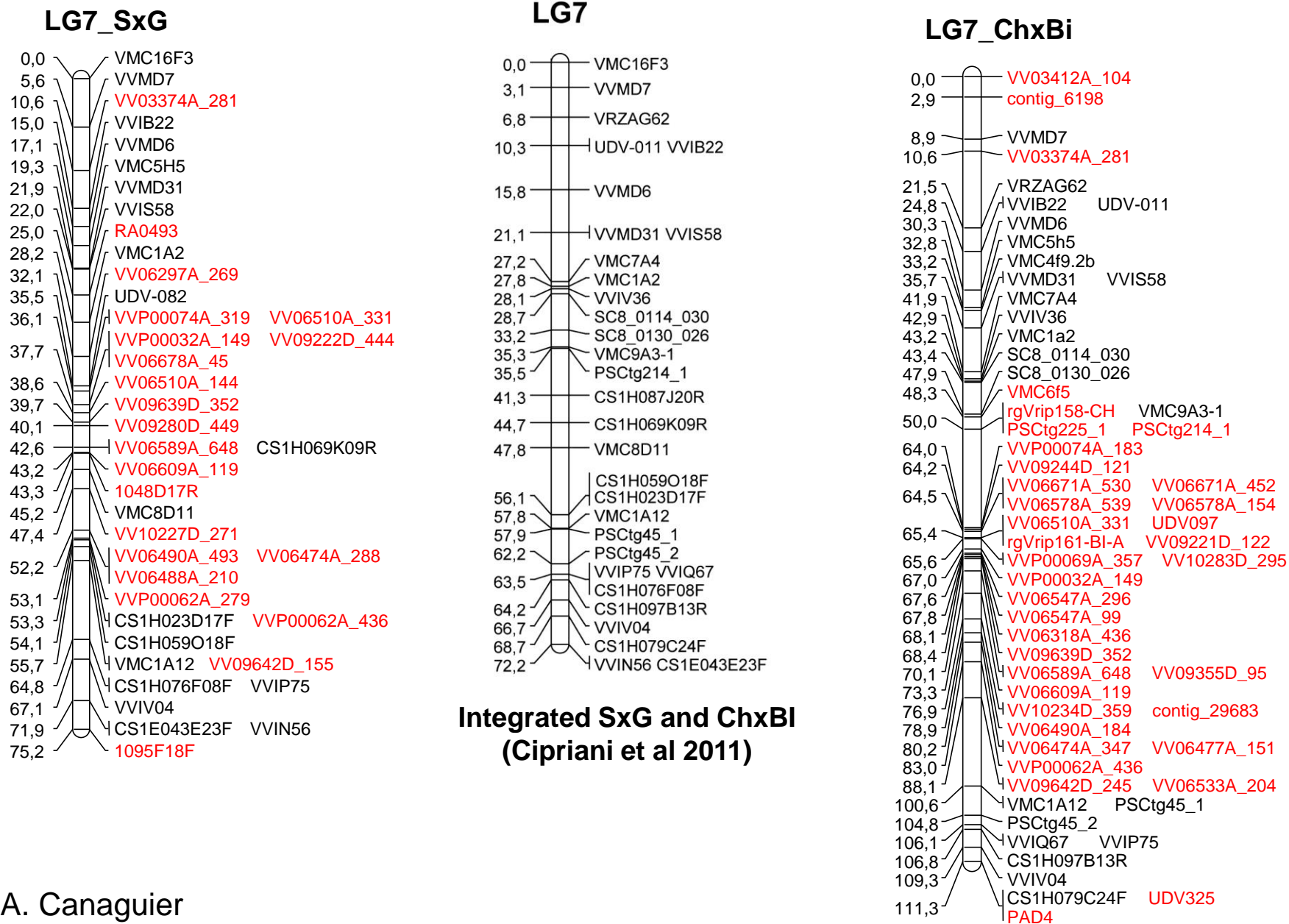
n° super-contigs



10.22% of 'N'

2064 gene models (7.8%)

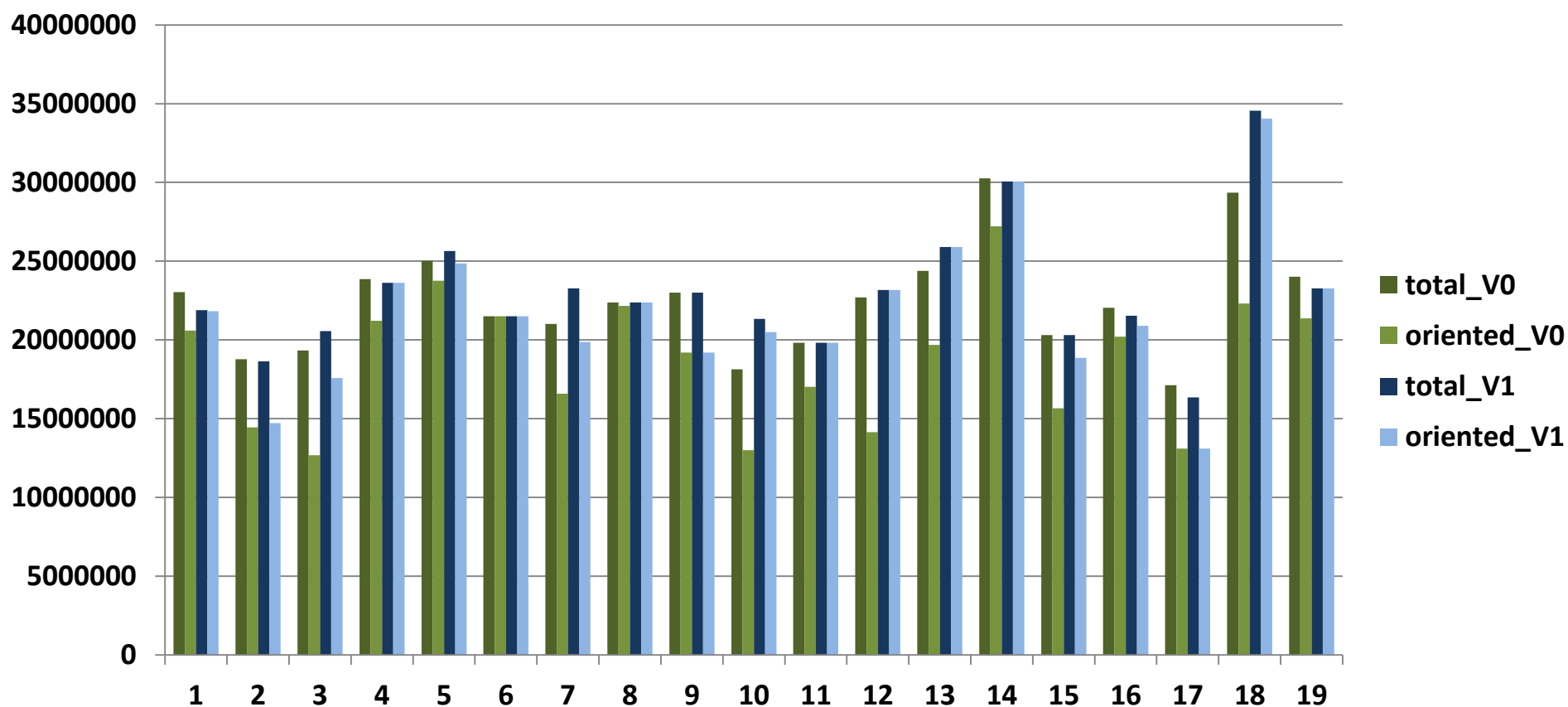
Improvement of the genetic map



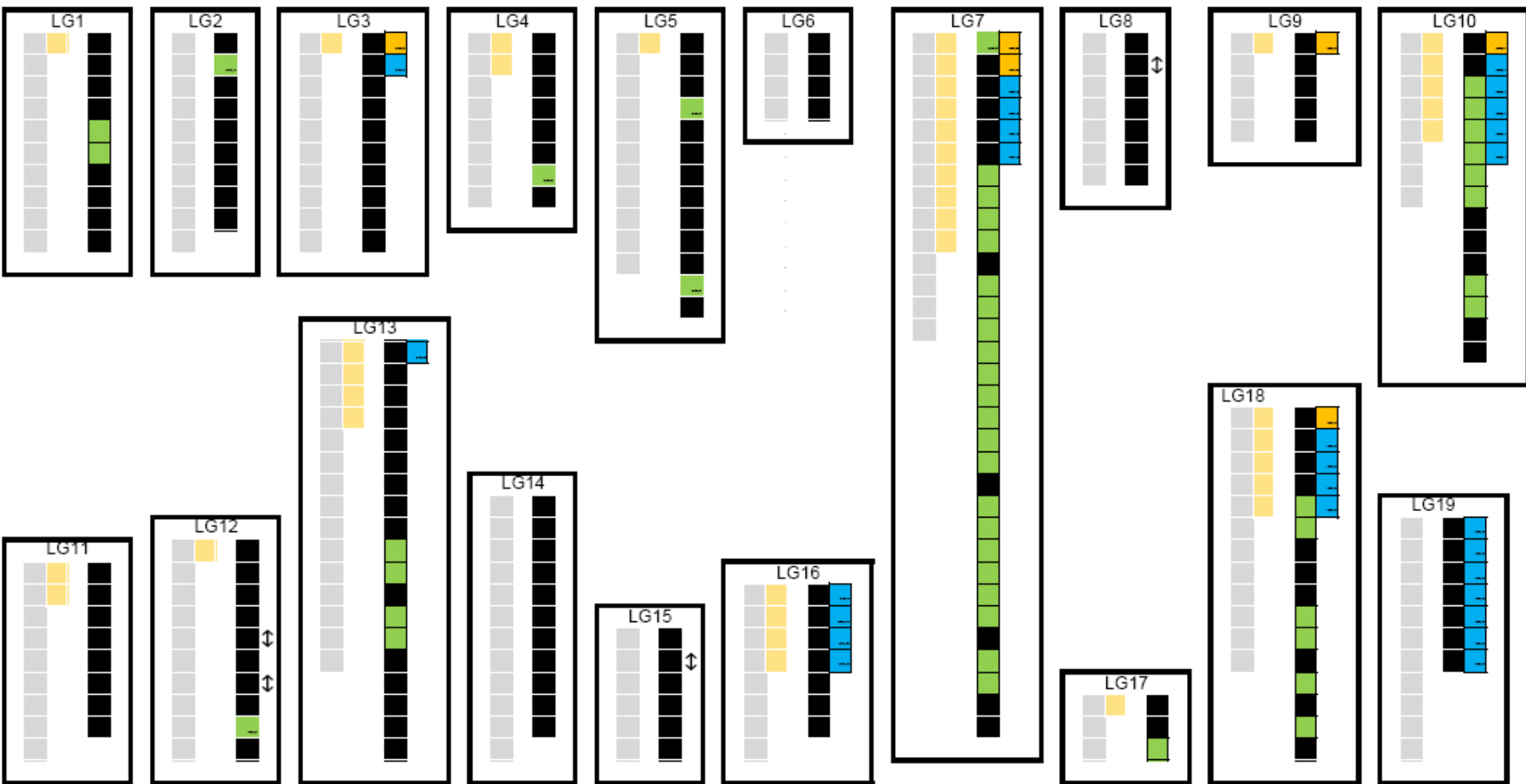
Anchoring the scaffolds using the new genetic map

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



Scaffold anchored and oriented V0
Scaffold anchored and oriented

New scaffold anchored and oriented

Scaffold with ambiguous location V0
Scaffold with ambiguous location V1

New scaffold with ambiguous location

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
n°chr

sYYYY
n°super-contig

gZZZZ
n°gene on super-contig

Improvement of the V1 annotation

- 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)
 - 48 genes/pseudogenes of the CHS and stilbene synthase gene family (Vannozzi et al 2012; Parage et al 2012; 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

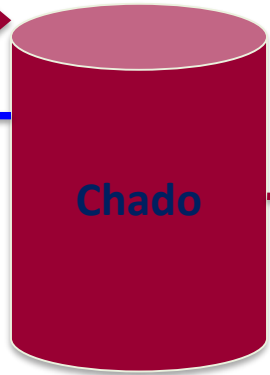
<http://urgi.versailles.inra.fr/index.php/urgi/Species/Vitis/Resources/Genome-sequence-and-annotations>

Apollo (login required)



Apollo WebStart

Write



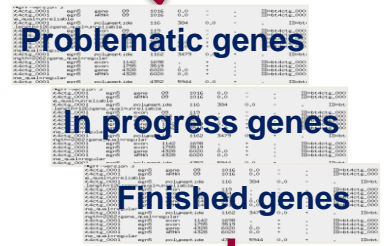
Read



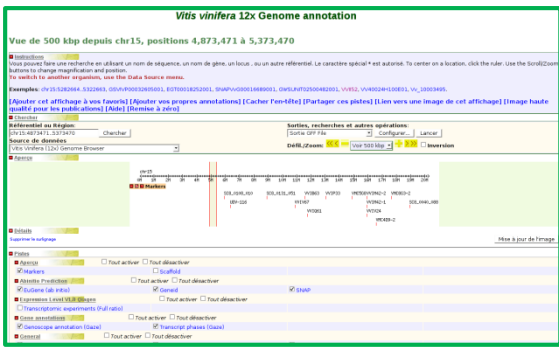
Export Edited genes (GFF3)



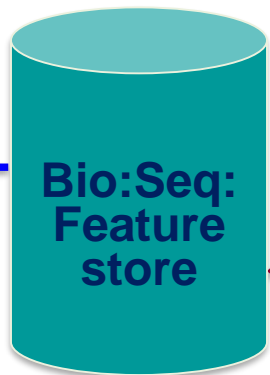
Never used! Too complex?



Gbrowse update



Read



New complete release



Improvement of the V1 annotation

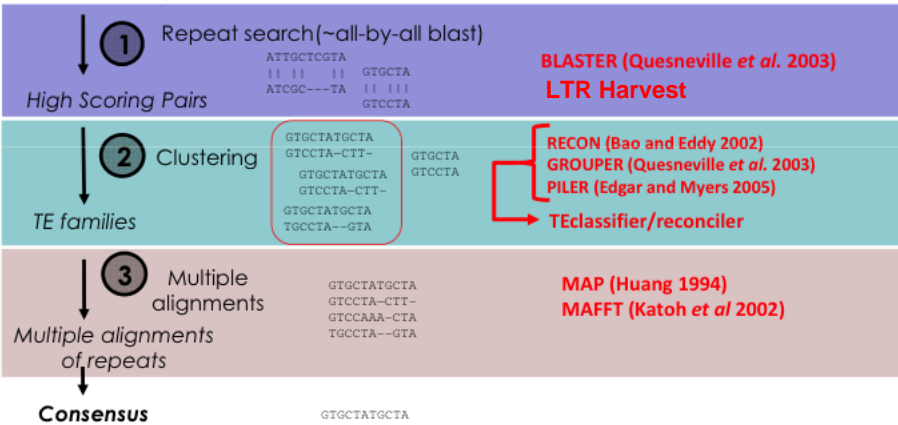
- Automatic annotation by IGA (Udine, Italy)
- Curation of the Class II transposons (Benjak et al 2008, 2009)

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

TE identification (de novo)

Genomic sequence ...TATGTGCTATTACTATTACATTACCATGCGT...

Pipeline TEde novo
(Flutre et al. In prep.)



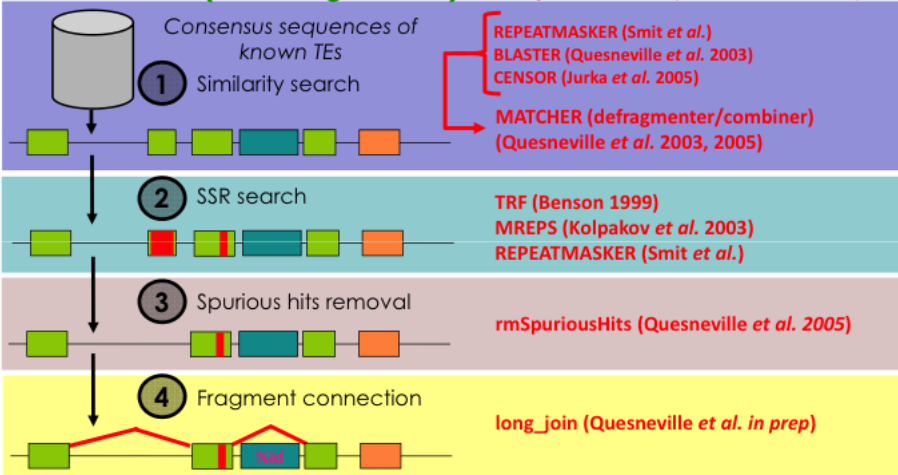
Novel TE automatic annotation using REPET V2 (Flutre et al 2011)

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

TE annotation (knowledge based)

Pipeline TEannot (Quesneville et al. 2005)



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 GnpIS portal

Access to Vitis resources

The screenshot shows the URGi website homepage. At the top left is the URGi logo. The main header reads "PLANT AND FUNGI DATA INTEGRATION". A navigation bar includes "Platform", "Research", "Projects", "Data", "Tools", and "Species". A search bar is on the right. The main content area features a large "GNPIS PORTAL" section with a circular diagram of tools (GnpMap, GnpSeq, GnpGenome, GnpArray, GnpPro, GnpSup, SIREgal, Ethesis) and a "SPECIES" section with a corn image. A "RESEARCH" section is at the bottom. A "WHAT'S NEW?" sidebar on the right lists updates for GnpIS 1.6.1, Siregal 1.6.1, and REPET 1.3.12. Two red arrows point from text boxes above to the GnpIS portal diagram and the Vitis resources section.

FEEDBACK | CONTACT | SITE MAP | ABOUT US

Register

Platform Research Projects Data Tools Species

SEARCH [input] OK

WHAT'S NEW?

17 Nov 2010 **GnpIS portal 1.6.1**
[GnpIS 1.6.1](#) is now available.

20 Oct 2010 **Siregal 1.6.1**
[Siregal 1.6.1](#) is now available:
• add basket tool
• add accessions ordering

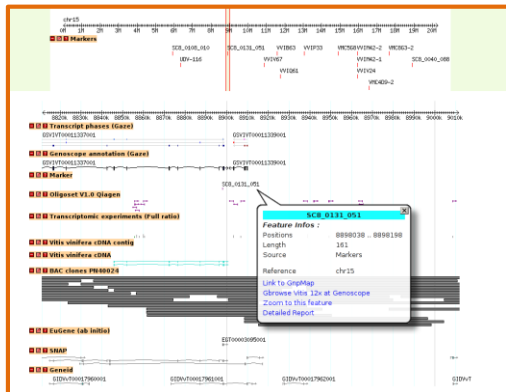
08 Oct 2010 **REPET 1.3.12**
[REPET 1.3.12](#) is now available.

29 Sep 2010 **SRS**
EMBL (Divisions Plant and Fungi)
Release 105 is now available on SRS

07 Sep 2010 Development of a workflow for

02 Dec 2010 Siregal and Ethesis steering

Markers



Vitis web pages

Genome sequence and annotations

- The [grapevine genome 12X draft sequence](#) and its annotations are available with GBrowse
- Access to the private annotations of the grapevine genome 12X sequence
- Download the genome sequence and annotations from the 12X genome browser
- [CMap 12X and 8X](#): Comparative map tool: Genome sequence links with the grapevine reference genetic map.

The [grapevine genome 8X draft sequence](#) and its annotations are available with GBrowse. The retrieve and annotation of gene families is facilitated in [FLAGdb++](#).

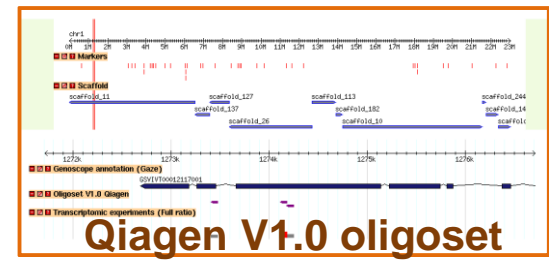
Apollo

Distributed manual annotation system

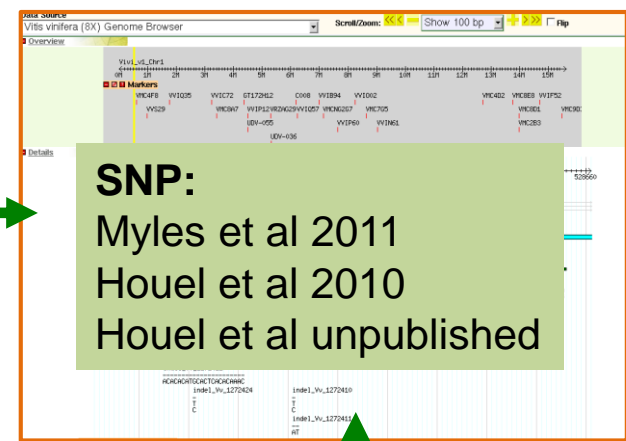
Apollo Webstart Under the supervision of the Annotation Committee Program, URGI provides tools for the manual curation of the structural version.

The access is restricted to the grapevine community. Please [contact us](#).

Transcriptomics data (GRASP project)



Polymorphism data



Map db

27 Vitis maps
New: Integrated_2

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

12X.0 GBrowser

Vitis vinifera 12x Genome annotation

Vue de 500 kbp depuis chr15, positions 4,873,471 à 5,373,470

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

SIREGAL

Genetic resources db

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

GnpSNP

Polymorphism db

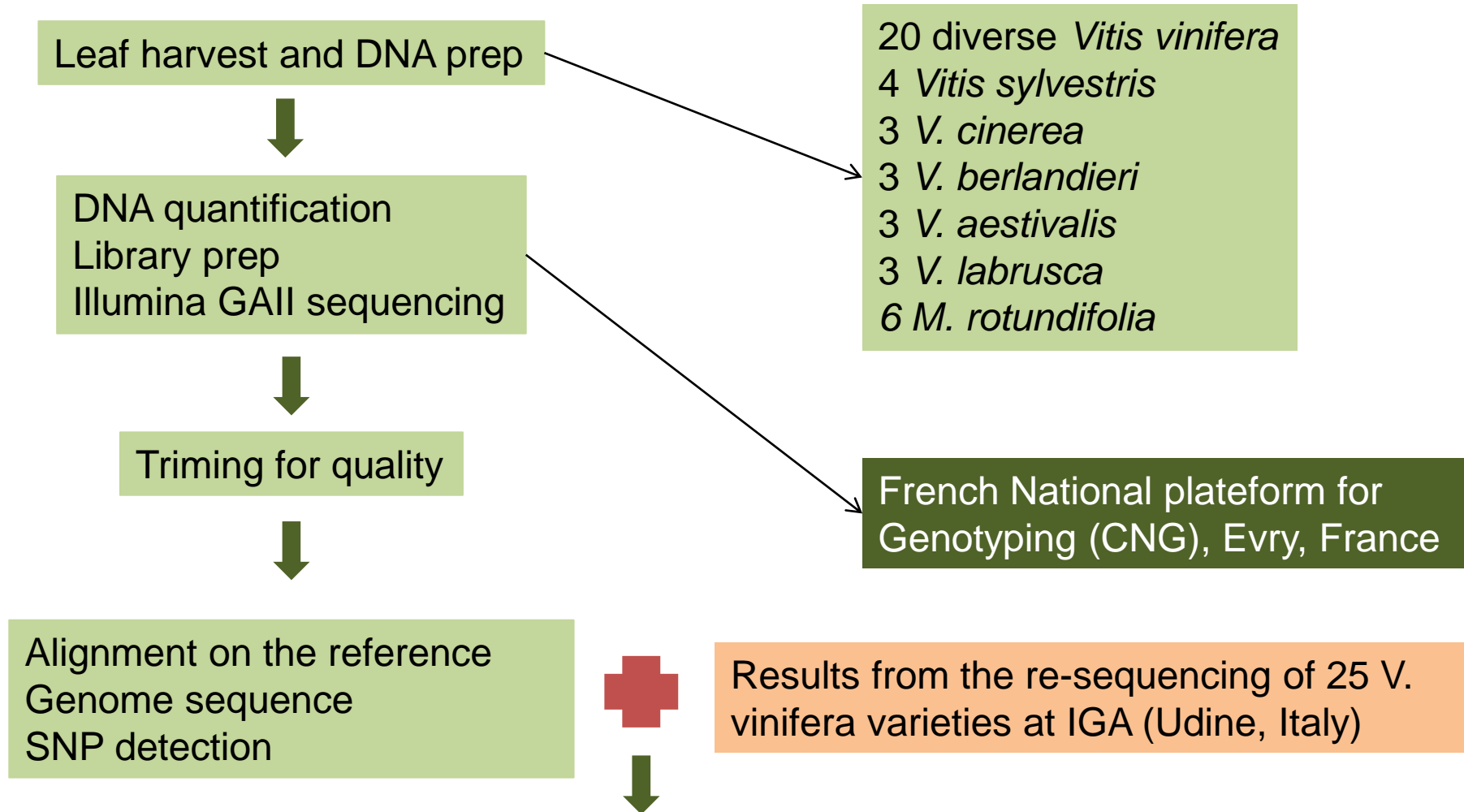
> 1.6 million variants

Line	SNP	Position
Line	50:109304210	243 / 218
Line	9157Map1Schwarzspann	236 / 218
Line	9221Map1Schwarzspann	243 / 214
Line	9500Map13Collette II	236 / 239
Line	308Map1Muscat of Alexandria B	243 / 287
Line	643Map14Mauzac B	227 / 233
Line	193Map1Pivot noir N	235 / 216

Molecular basis of trait variation
and development of MAS at a
genomic scale

Illumina re-sequencing of 47 grapevine genotypes

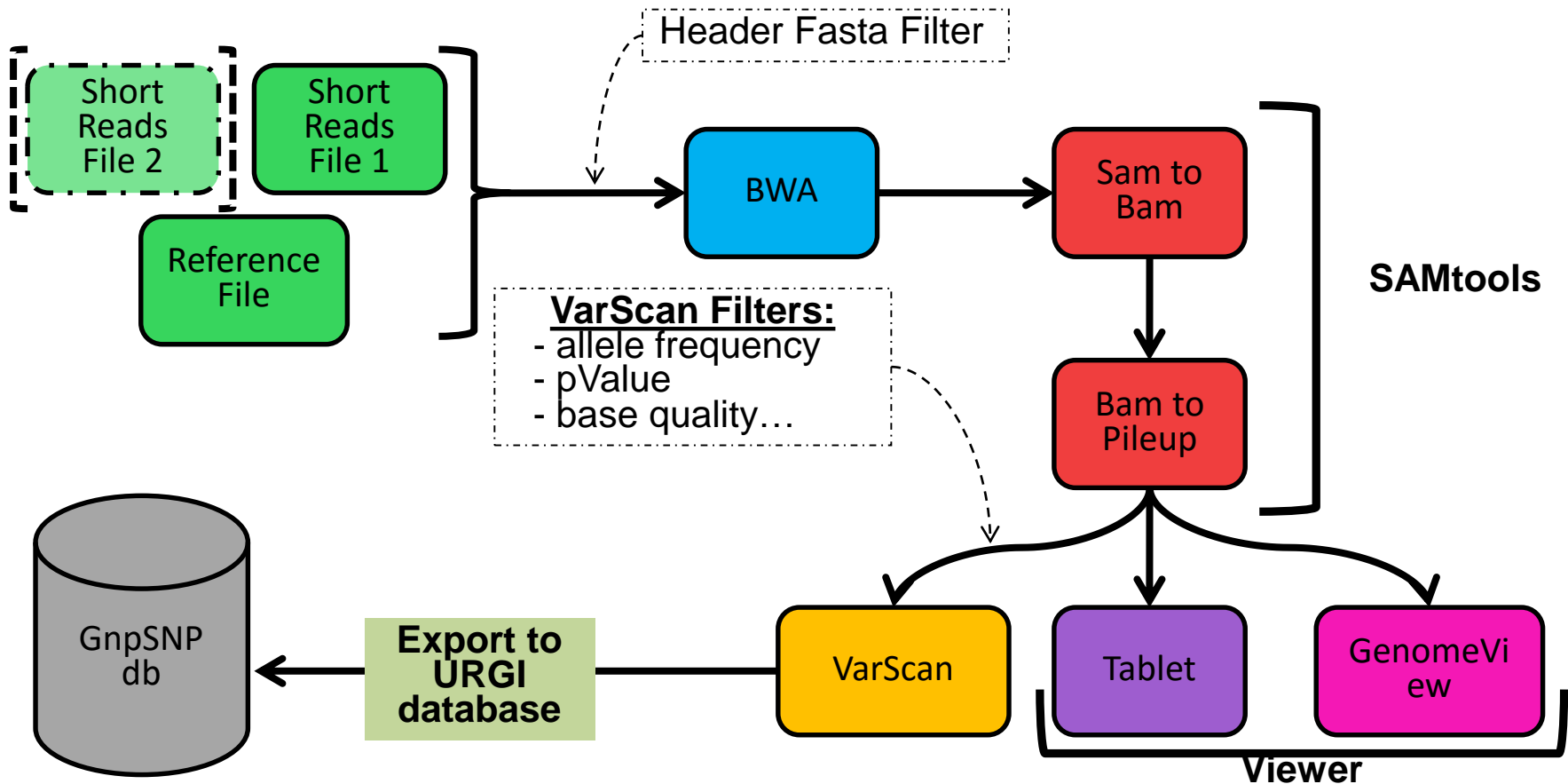
In the frame of EU KBBE GrapeReSeq project and of a french ANR project Muscares



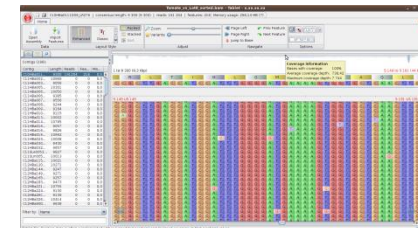
Illumina 20K genotyping Chip (<http://urgi.versailles.inra.fr> => Vitis pages)

Pipeline for SNP detection: MAPHITS

Mapping Analysis Pipeline for High-Throughput Sequencing



Calculations parallelized



MAPHiTS workflow in Galaxy

(a web-based platform for genomic research : <http://usegalaxy.org>)

Galaxy Analyze Data Workflow Data Libraries Help User

Tools

- GraphDisplay Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- FASTA manipulation
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: Peak Calling
- Rg Data
- Rg Simulate
- Rg Visualise
- Rg Model Data
- URGI NGS
 - Map with Bwa 0.5.7 Map Short Reads to Reference sequence with BWA 0.5.7.
 - Map with Bwa 0.5.7 (cmd lines) Map Short Reads to Reference sequence with BWA 0.5.7 (and command lines).
 - Header Fasta Filter Modify all of this header file that contains one, or multiple, tabulation and informations after the name of the sequence.
 - Sam Filter All alignments that are outside the subject sequence (partially or not), are removed from the input file and copy to an output file.
 - VarScan VarScan: convert bam file to pileup file with differents filters.
 - Tablet (v.1.10.03.04) Alignment Viewer (can use GFF3, ACE, AFG, MAQ, SOAP, SAM and BAM Files).
 - GenomeView (v.922) Alignment Viewer (can use BAM, GFF, FASTA and Annotation Files).
- URGI Mapping BES

Workflow Canvas | Unnamed workflow Options

Input Dataset → Map with Bwa 0.5.7 → SAM-to-BAM → Generate pileup → Tablet (v.1.10.03.04)

Details

Tool: Tablet (v.1.10.03.04)

Fichier Bam / Sam / Map / autres...
Data input 'input_FICHIER' (bam)

RÃ©fÃ©rence Input.fasta
Data input 'input_REF' (fasta)

Edit Step Attributes

Annotation / Notes:

Add an annotation or notes to this step; annotations are available when a workflow is viewed.

Tablet:

Tablet is a lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments. This software is able to view sequences alignments. It can use GFF3, ACE, AFG, MAQ, SOAP, SAM and BAM Files.

Tablet features:

- * High-performance visualization and data navigation.
- * Display of reads in both packed and stacked formats.
- * File format support for ACE, AFG, MAQ, SOAP, SAM and BAM.
- * Import GFF3 features and quickly find/highlight them.
- * Search and locate reads by name across entire data sets.
- * Entire-contig overviews, showing data layout or coverage.
- * Simple install routine via auto-updating graphical install.
- * Support for Windows, Apple Mac OS X, Linux and Solaris, i

Website: <http://bioinf.scri.ac.uk/tablet/>

<http://localhost:8080/workflow/editor?id=f06d708dacad69d8#>

Sequencing results for the 47 *V. vinifera* varieties

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

- GA II and/or High Seq sequencing
- Filtering for position in known structural variants, in repeats, with too few or too many reads



20k Illumina GrapeReSeq genotyping chip

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 ≥ 0.9
- SNP for Infinium type II array
- MAF criteria:
 - 90% SNP with $MAF > 0.1$ (85463 SNPs)
 - 10% SNP with $0.05 < MAF < 0.1$ (27631 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. Weissahr

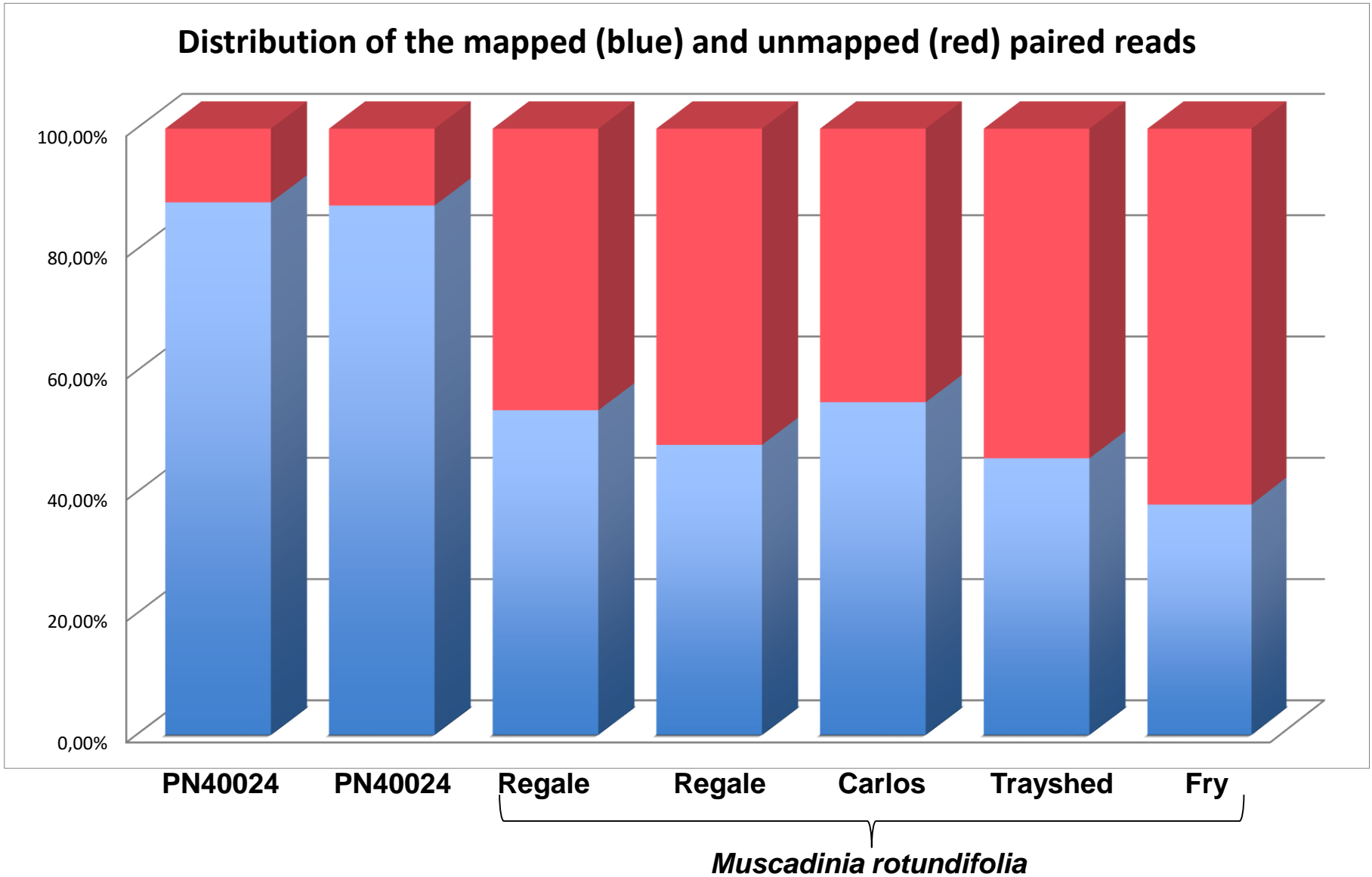


CIVV, Spain
J-M Martinez-Zapater

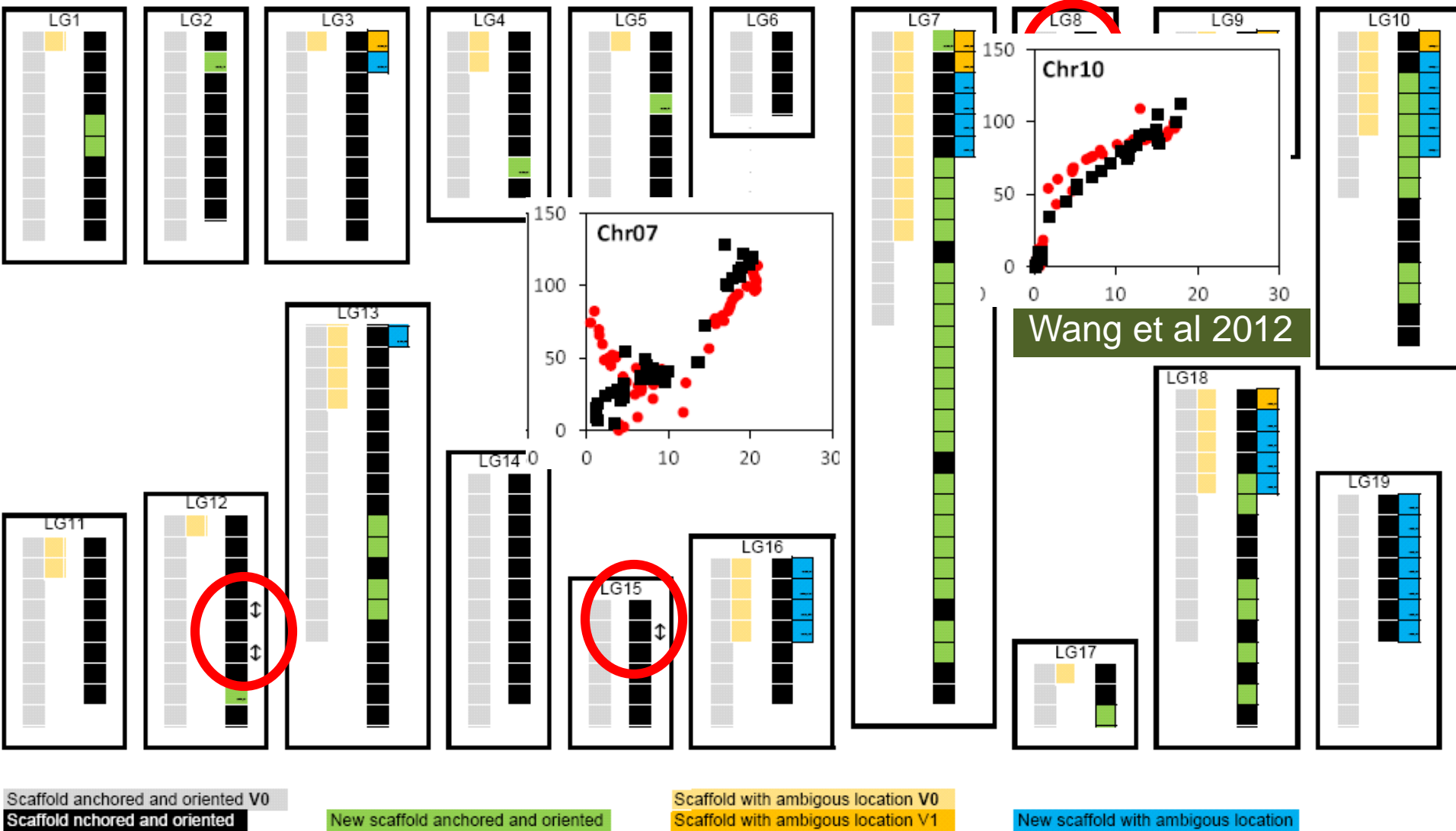
Thank you!



Mapping with 99% of identity



Anchoring the scaffolds using the new genetic map



Nearly ready to provide a new chromosome sequence

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)

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