



Uncovering the wealth of grapevine genetic diversity through whole genome sequencing and assembly

Dario Cantù

Associate Professor and Louis P. Martini Endowed Chair



VITICULTURE & ENOLOGY

COLLEGE OF AGRICULTURAL
AND ENVIRONMENTAL SCIENCES



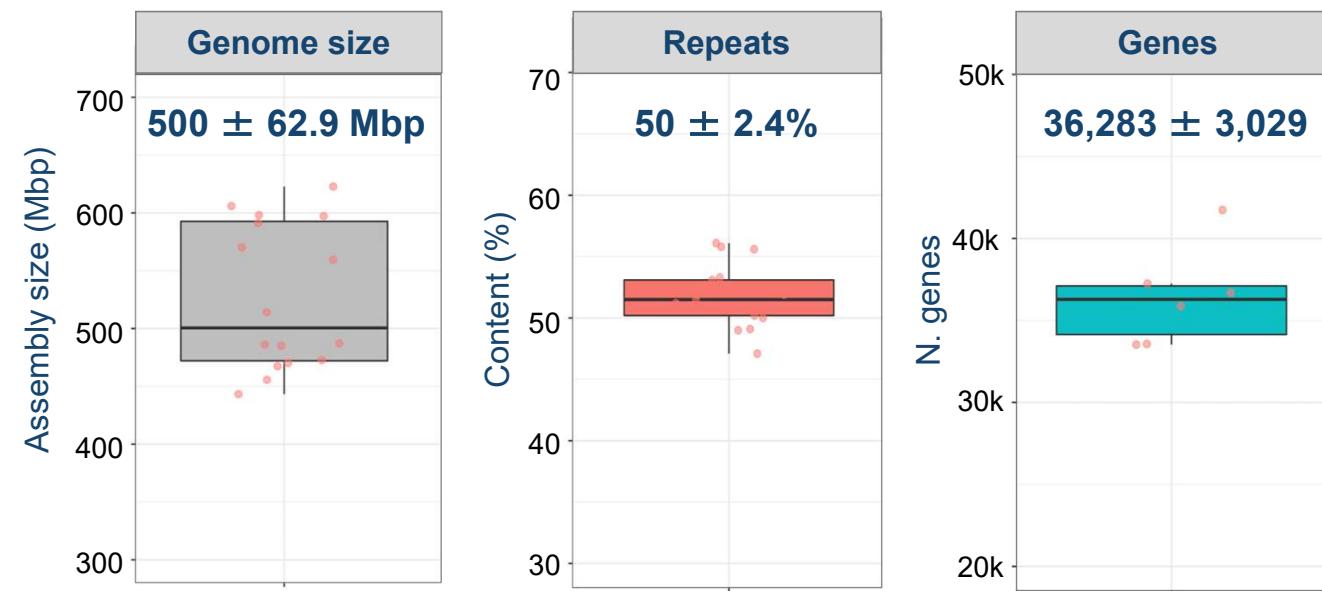
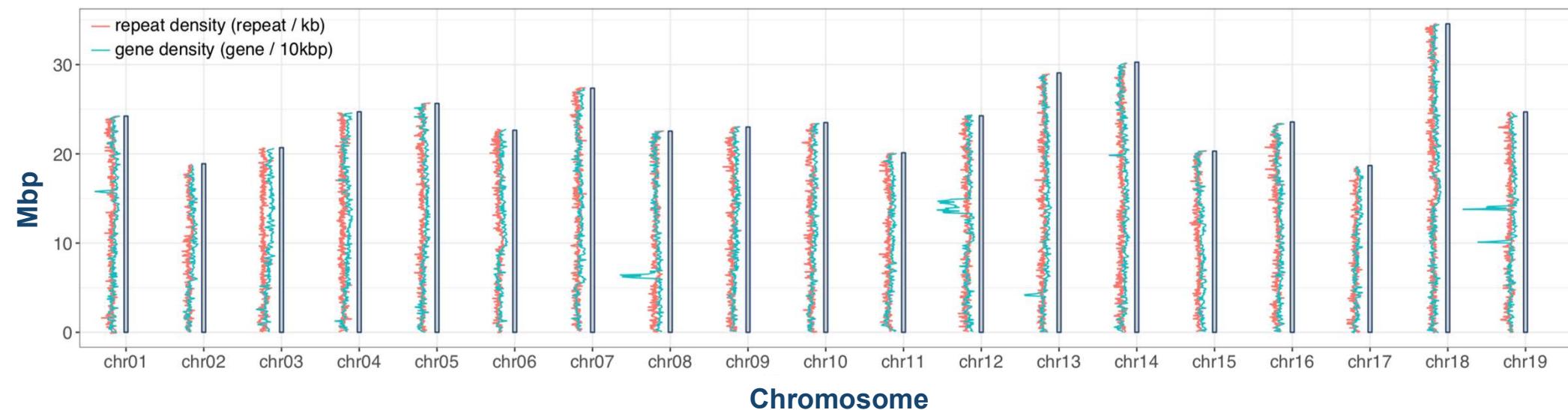
Outline

1. Why do we need more genome references?
2. What do we need to generate and study more (high quality) grape genomes?
3. Where are we now and where are we going?
4. What do we still need?



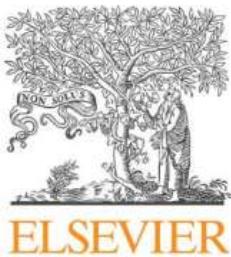
The grape genome

PN40024 12X V2 Canaguier et al., 2017



The first grapevine genomes (2007)

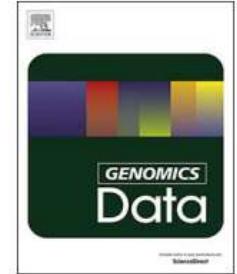
Genomics Data 14 (2017) 56–62



Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata



Data in Brief

A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3)



A. Canaguier^{a,b}, J. Grimplet^c, G. Di Gaspero^d, S. Scalabrin^d, E. Duchêne^e, N. Choisne^f, N. Mohellibi^f, C. Guichard^{a,g}, S. Rombauts^{h,i}, I. Le Clainche^{a,b}, A. Bérard^b, A. Chauveau^b, R. Bounon^{a,b}, C. Rustenholz^e, M. Morgante^d, M.-C. Le Paslier^b, D. Brunel^b, A.-F. Adam-Blondon^{a,f,*}

^a UMR GV, INRA, UEVE, ERL CNRS, 2 rue Gaston Crémieux, 91000 Evry, France

^b EPGV US 1279, INRA, CEA, IG-CNG, Université Paris-Saclay, 91000 Evry, France

^c Instituto de Ciencias de la Vid y del Vino (CSIC, Universidad de La Rioja, Gobierno de La Rioja), Logroño 26007, Spain

^d IGA, via J. Linussio 51, 33100 Udine, Italy

^e SVQV, UMR 1131, INRA, Université de Strasbourg, 28 rue de Herrlisheim, 68000 Colmar, France

^f URGI, UR 1164, INRA, Université Paris-Saclay, route de Saint-Cyr, 78026 Versailles, France

^g IPS2, UMR 1403, INRA, Université Paris-Saclay, Rue de Noetzlin, bât. 630, 91190 Gif-sur-Yvette, France

^h Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 927, 9052 Ghent, Belgium

ⁱ VIB Center for Plant Systems Biology, Technologiepark 927, 9052 Ghent, Belgium

1. Why do we need more grape genome references?

Di Genova et al. BMC Plant Biology 2014, 14:7
<http://www.biomedcentral.com/1471-2229/14/7>



RESEARCH ARTICLE

Open Access

Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants

Alex Di Genova^{1,2}, Andrea Miyasaka Almeida^{1,4}, Claudia Muñoz-Espinoza^{1,4}, Paula Vizoso^{1,4}, Dante Traversany^{1,2}, Carol Moraga^{1,4}, Manuel Pinto⁵, Patricio Hinrichsen^{5†}, Ariel Orellana^{1,4†} and Alejandro Maass^{1,2,3*}+
+✉

Venturini et al. BMC Genomics 2013, 14:41
<http://www.biomedcentral.com/1471-2164/14/41>

RESEARCH ARTICLE

Open Access

De novo transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity

Luca Venturini[†], Alberto Ferrarini[†], Sara Zenoni, Giovanni Battista Tornielli, Marianna Fasoli, Silvia Dal Santo, Andrea Minio, Genny Buson, Paola Tononi, Elisa Debora Zago, Gianpiero Zamperin, Diana Bellin, Mario Pezzotti and Massimo Delledonne^{*}

Thompson Seedless



Corvina

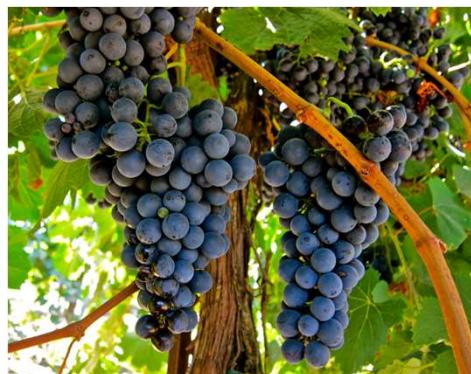


The High Polyphenol Content of Grapevine Cultivar Tannat Berries Is Conferred Primarily by Genes That Are Not Shared with the Reference Genome^W

Cecilia Da Silva,^{a,1} Gianpiero Zamperin,^{b,1} Alberto Ferrarini,^b Andrea Minio,^b Alessandra Dal Molin,^b Luca Venturini,^b Genny Buson,^b Paola Tononi,^b Carla Avanzato,^b Elisa Zago,^{b,2} Eduardo Boido,^c Eduardo Dellacassa,^c Carina Gaggero,^a Mario Pezzotti,^b Francisco Carrau,^c and Massimo Delledonne^{b,3}

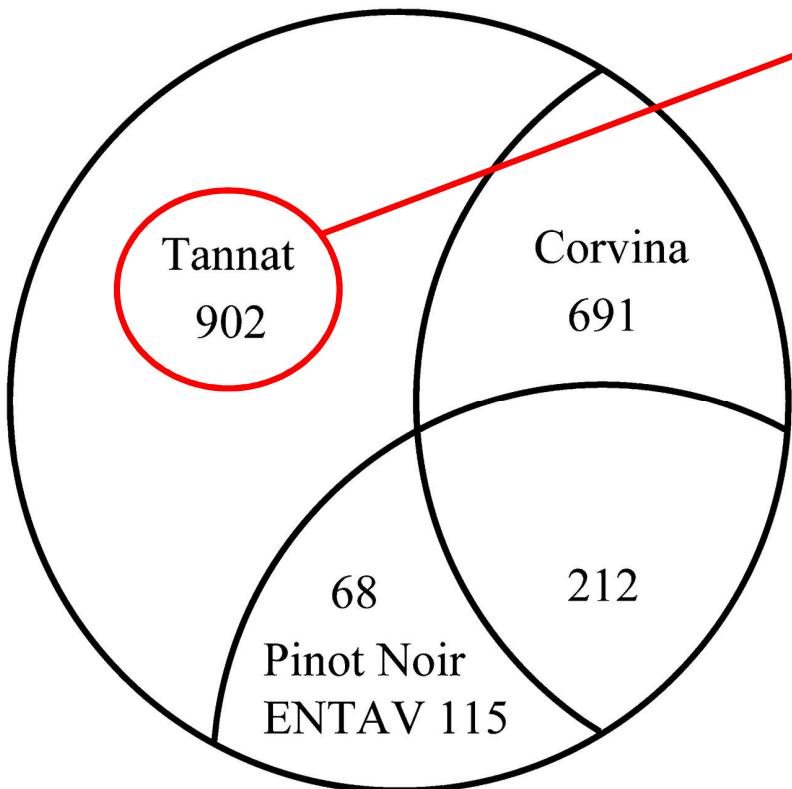
Tannat



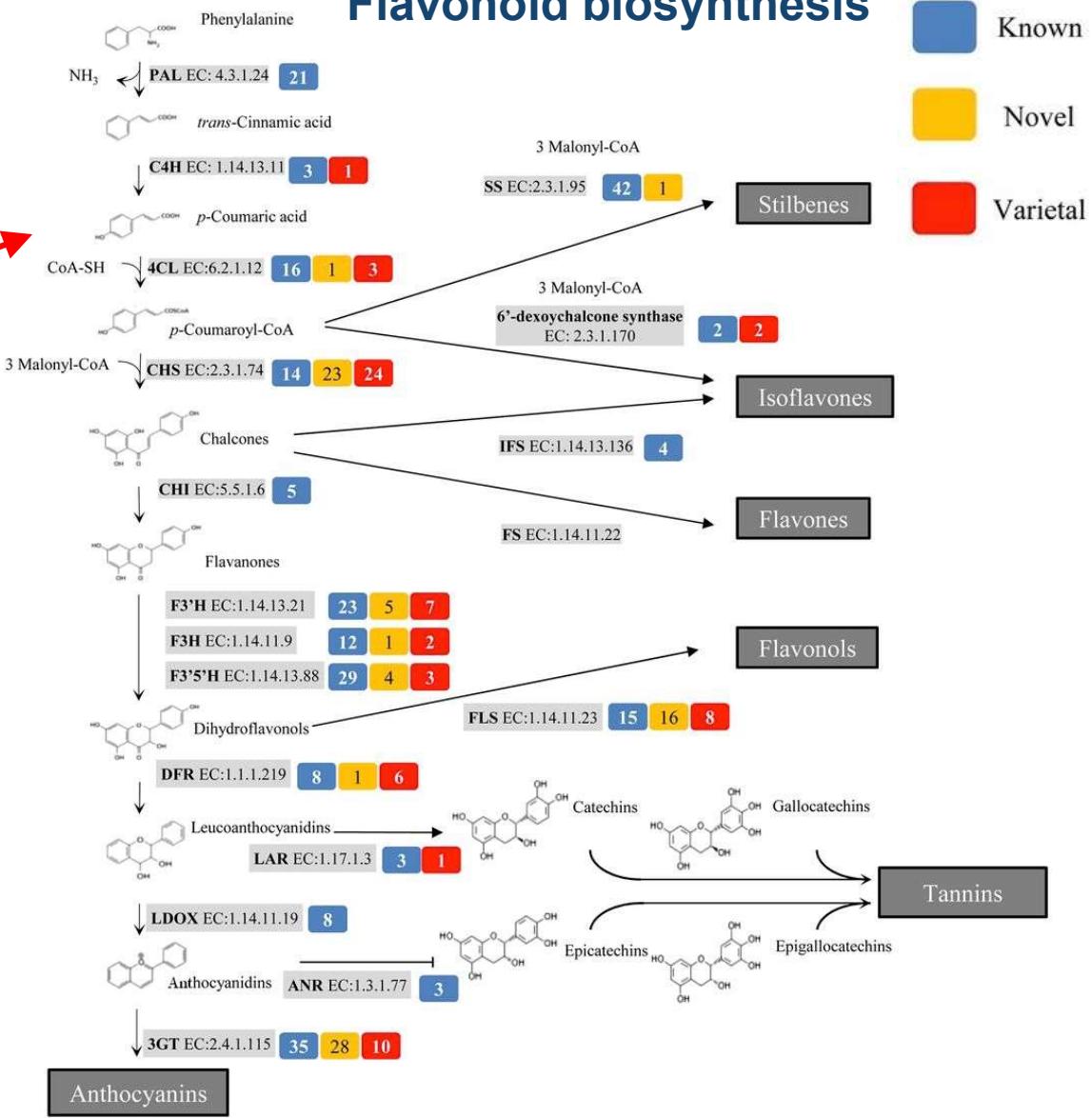


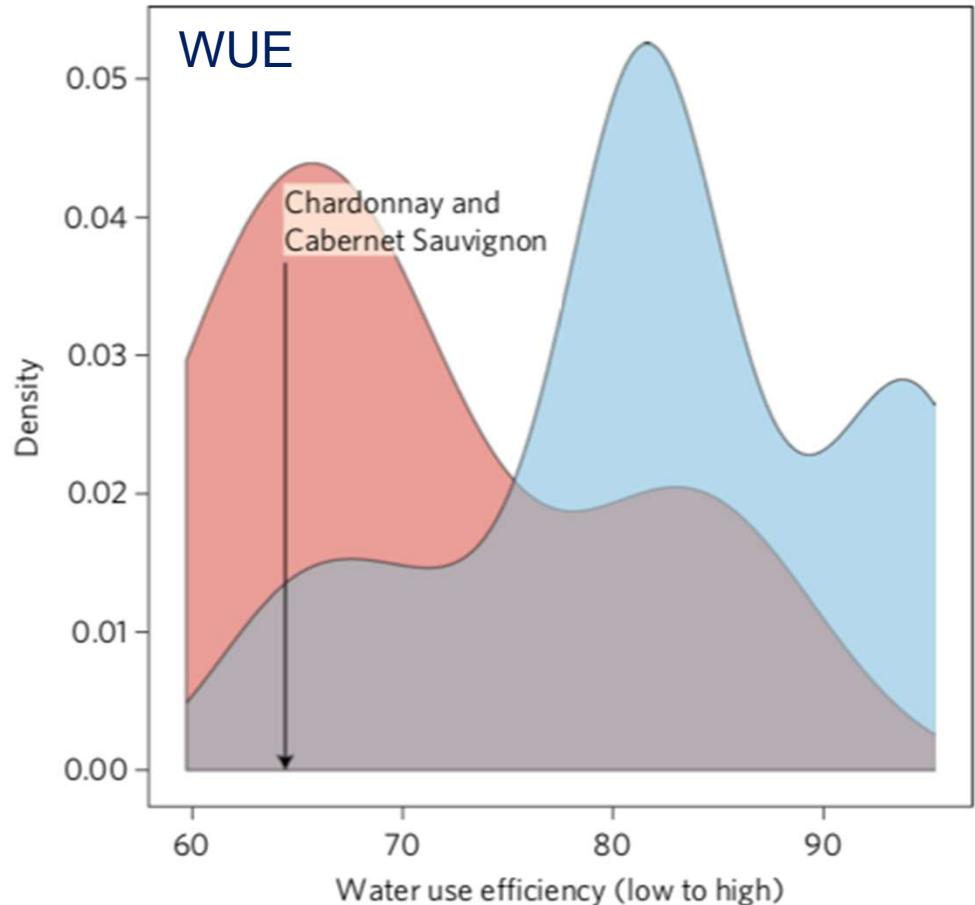
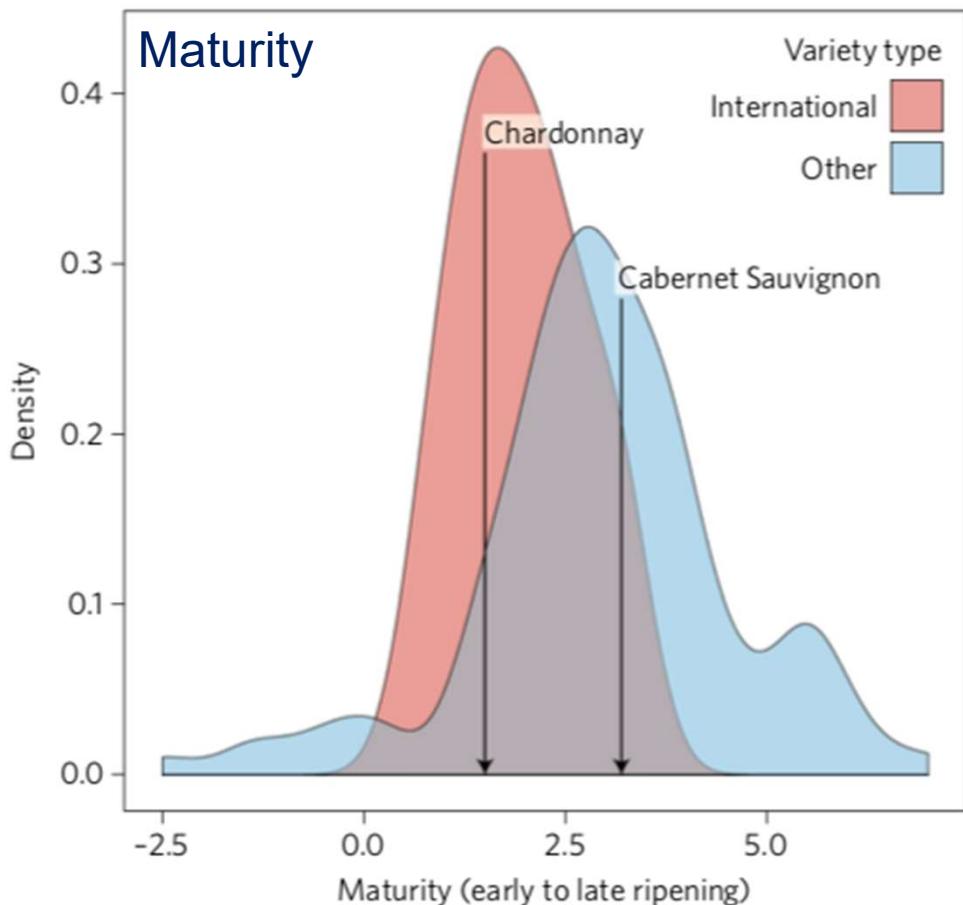
Categorization of the 1,873 genes not shared with PN40024. Number of genes found in common among Tannat, Corvina, and Pinot Noir (ENTAV 115)

V. vinifera
cv. Tannat

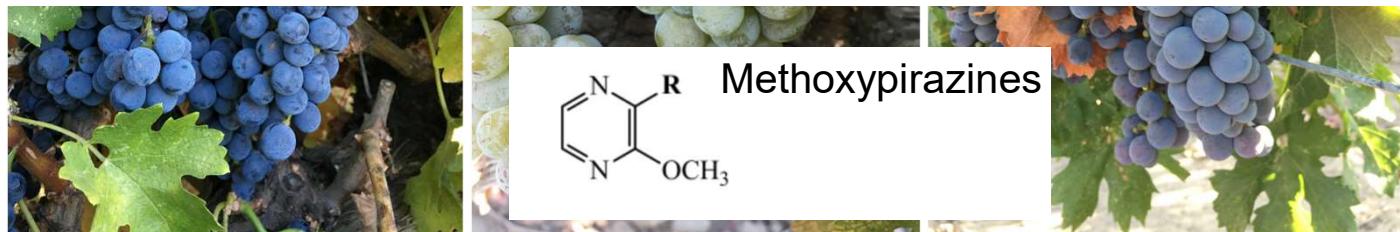


Flavonoid biosynthesis



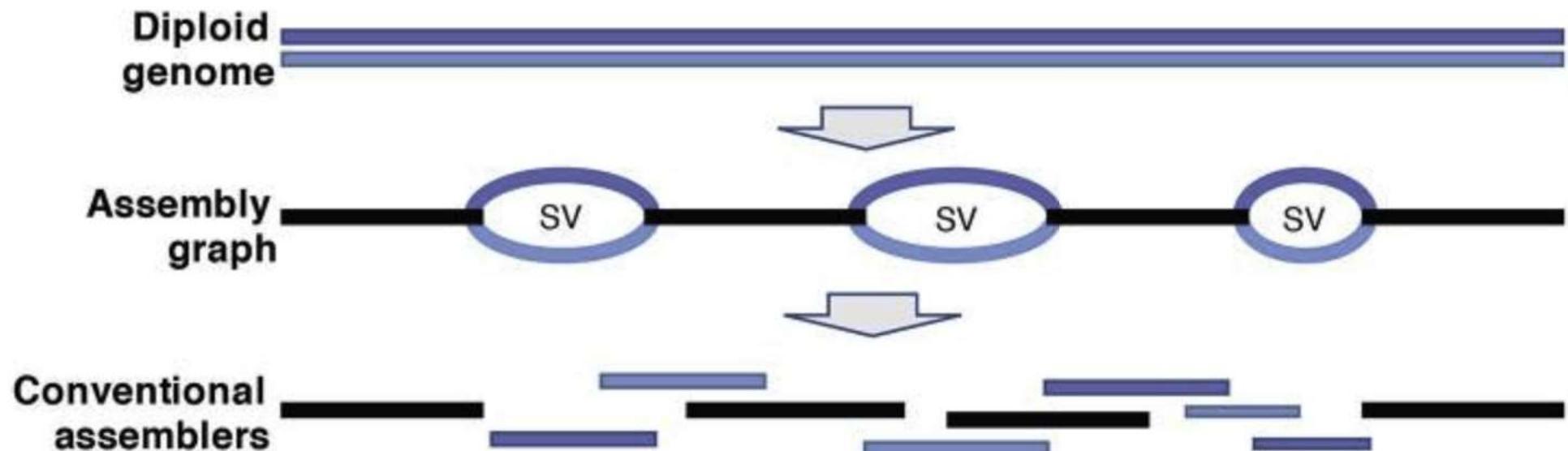


Wolkovich et al., 2017 Nature Climate Change



2. What do we need to generate and study more (high quality) grape genomes?

The challenge of heterozygosity



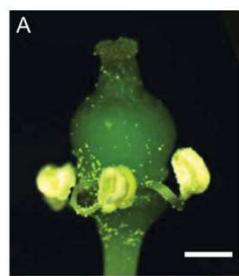
Minio et al., 2017



The challenge of heterozygosity

Wild grapes are dioecious

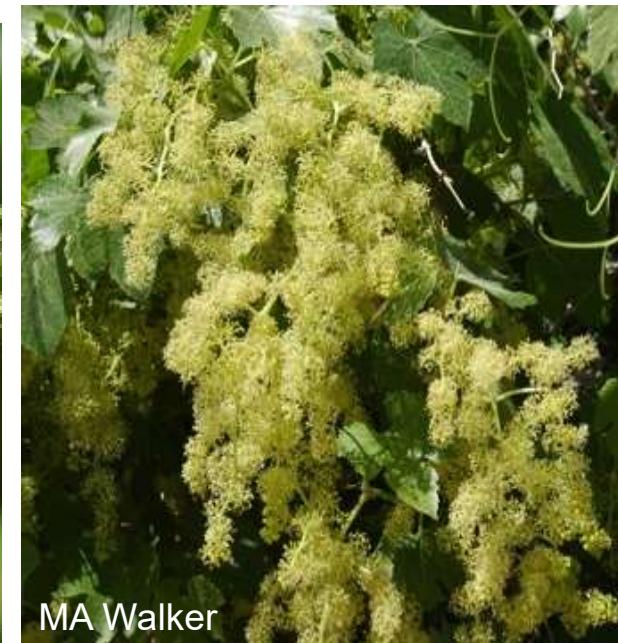
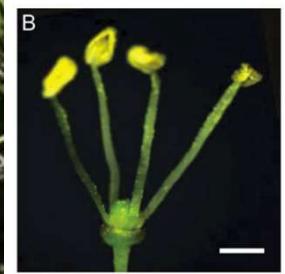
Female



Carmona et al., 2008



Male



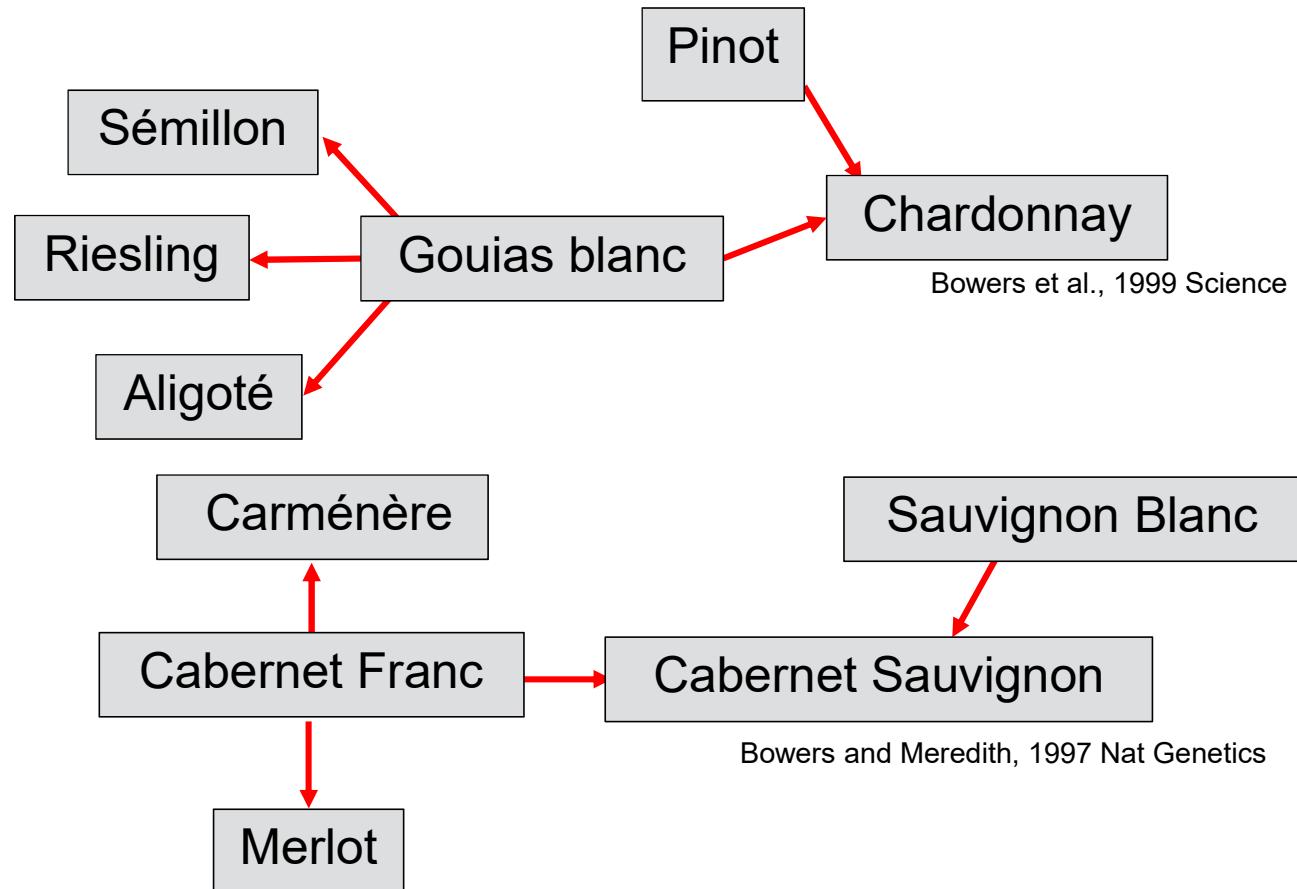
- Obligate out-crossers, so highly heterozygous
- High recessive load and strong inbreeding depression
- Hermaphrodites are rare in the wild since selfing expresses deleterious recessive alleles



Cultivated varieties are hermaphroditic, but suffer of inbreeding depression



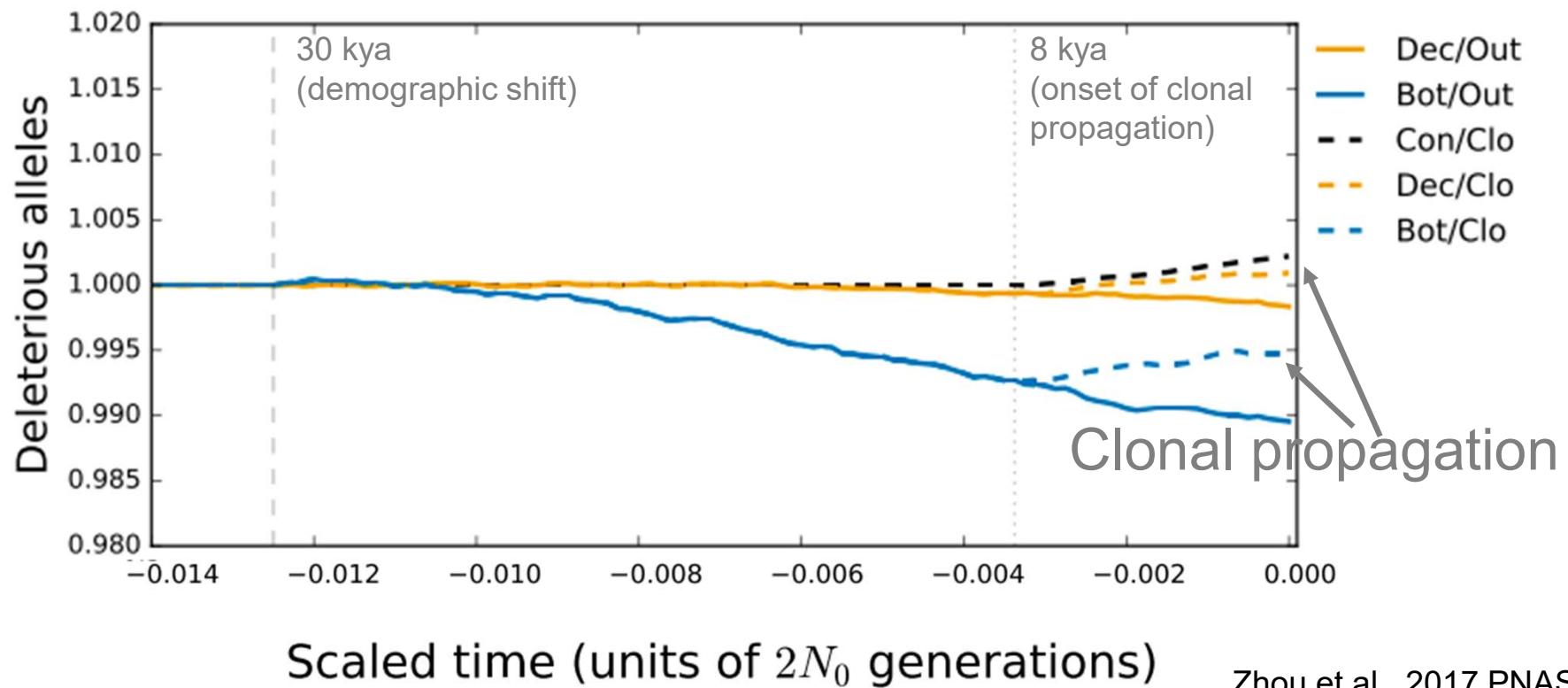
Examples of “famous crosses”





Cultivated varieties are hermaphroditic, but suffer of inbreeding depression

Forward simulations under a model of recessive selection for three demographic scenarios and two mating systems.





The challenge of heterozygosity

First attempts to
assemble
Cabernet
Sauvignon
(~2012)



Seq technology	Assembler	Assembly size	N. contigs	NG50	LG50
Illumina PE	SOAPdenovo2	631 Mb	994,414	2,325	49,584
Illumina PE	SPAdes	482 Mb	245,348	7,719	15,644
Illumina PE + PacBio (5x)	Celera	574 Mb	154,787	24,598	5,591

2. What do we need to generate and study more (high quality) grape genomes?

- a) **Better sequencing technologies:** longer reads (single molecule real time, nanopore), optical maps, and long-range scaffolding/phasing
- b) **Methods to extract pure and high-molecular weight DNA from grape tissue**
- c) **Assembly algorithms that enables the assembly of highly contiguous diploid genomes**

Cabernet Sauvignon FPS Clone 08



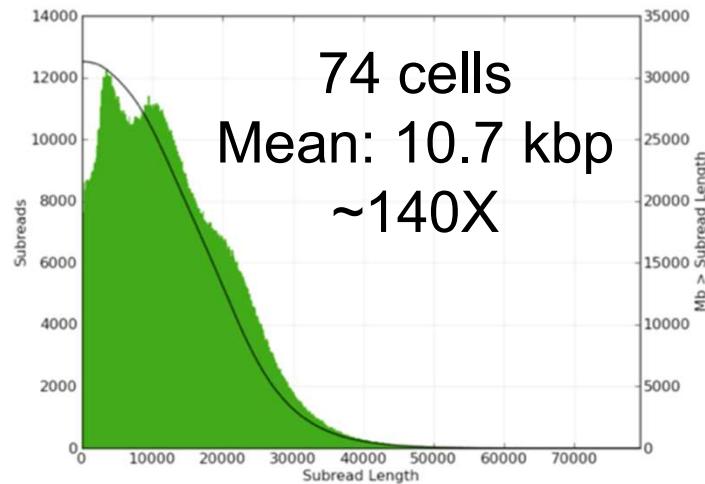
SMRT sequencing

20 kb & 30 kb
libraries



PACIFIC
BIOSCIENCES®

↓
P6-C4 chemistry
PacBio RSII



↓
FALCON-unzip
Quiver

Diploid contigs
(primary + haplotigs)

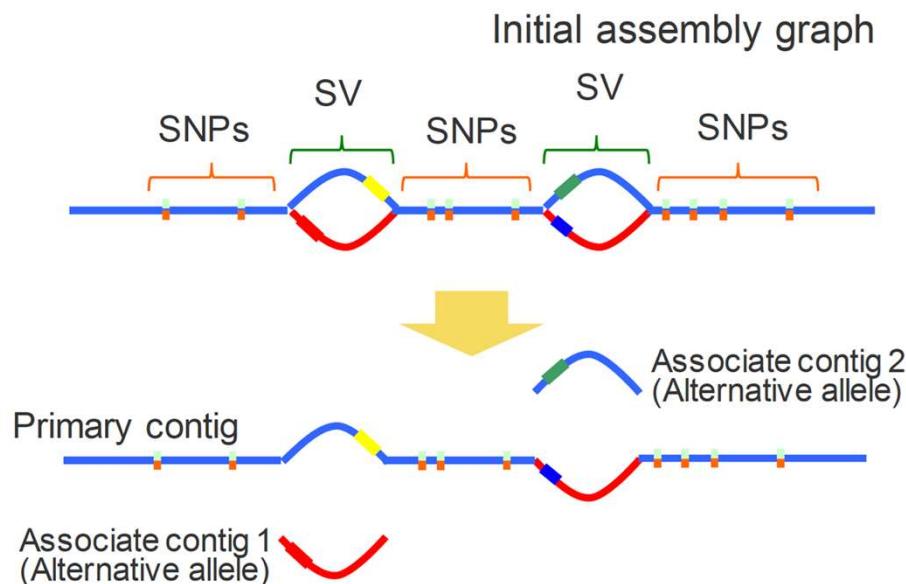
Common assemblers

SNPs

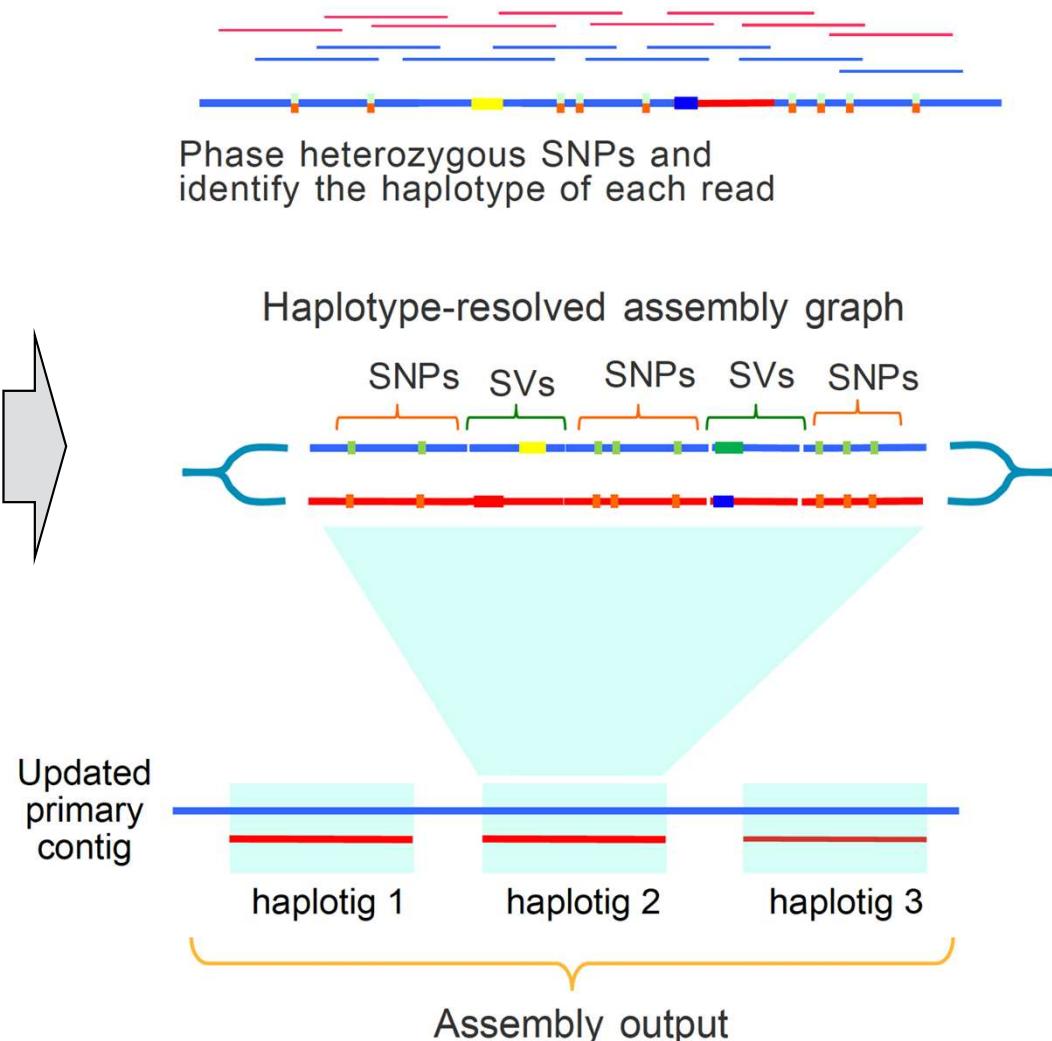
Contigs break at loci of
heterozygous structural variation

Haploid consensus

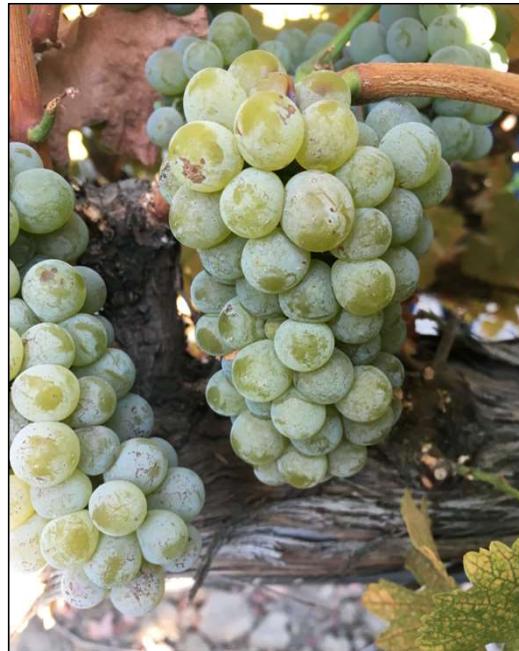
Falcon



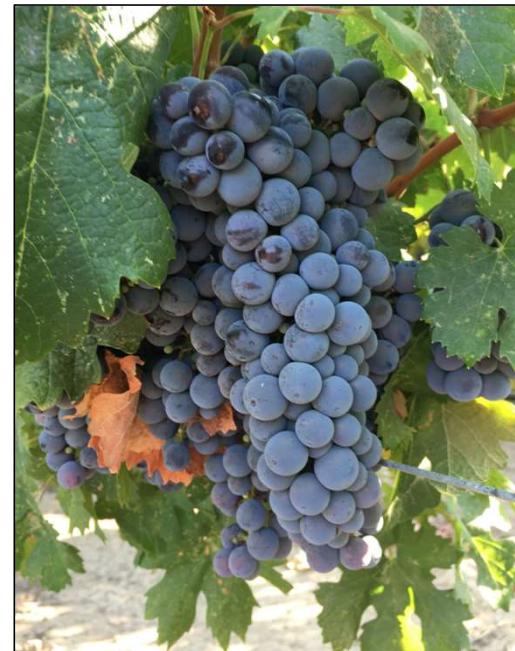
Falcon-unzip



*Sauvignon
Blanc*



X

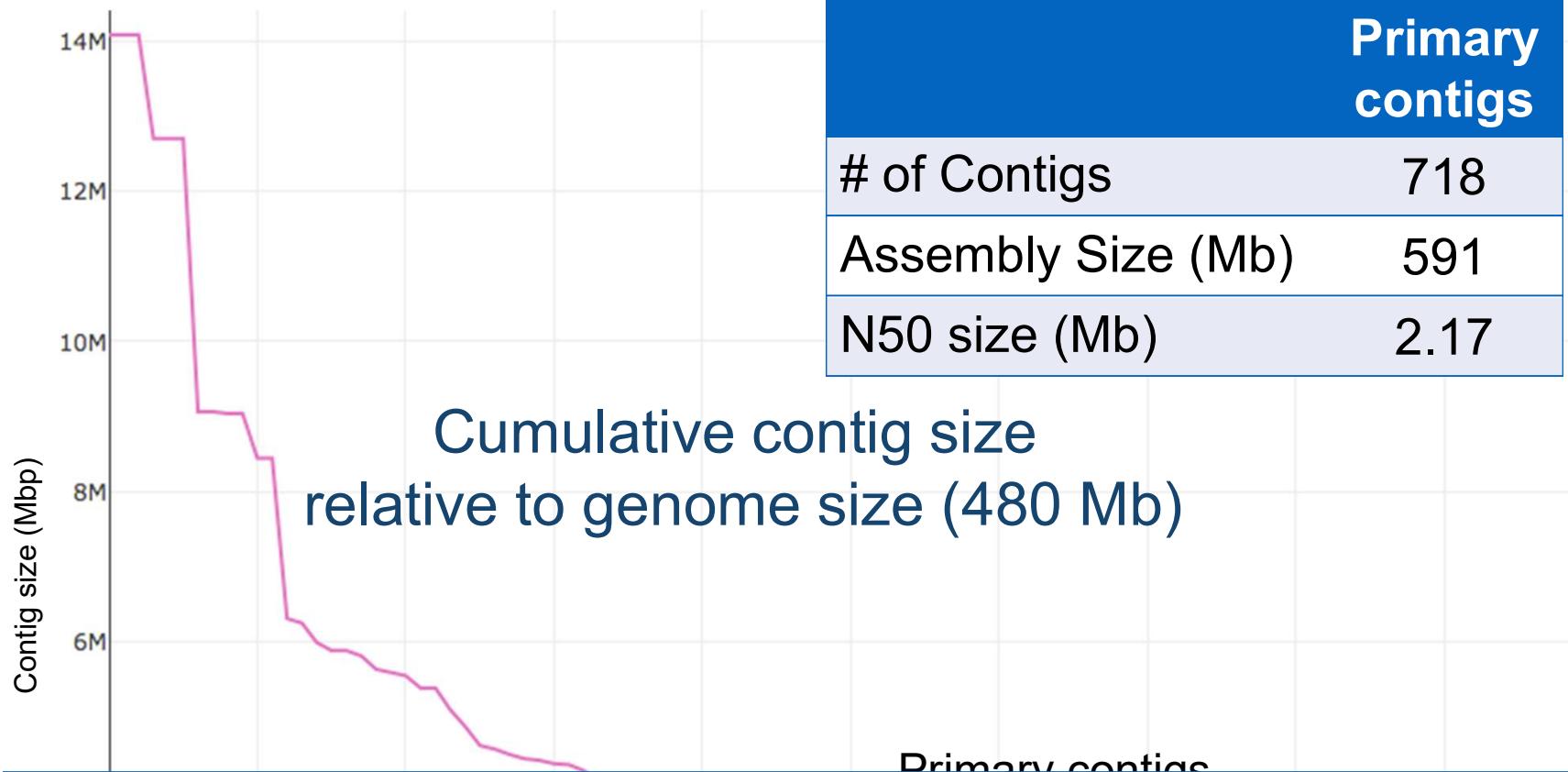


*Cabernet
Franc*



*Cabernet
Sauvignon*

Bowers and Meredith, Nat Genetics 1997



Seq technology	Assembler	Primary contigs			
		Assembly size	N. contigs	NG50	LG50
Illumina PE	SOAPdenovo2	631,320,289	994,414	2,325	49,584
Illumina PE	SPAdes	481,817,163	245,348	7,719	15,644
Illumina PE + PacBio (5x)	Celera	573,589,710	154,787	24,598	5,591
PacBio (20 + 30 kb lib)	FALCON-unzip	590,964,935	718	2,767,687	53

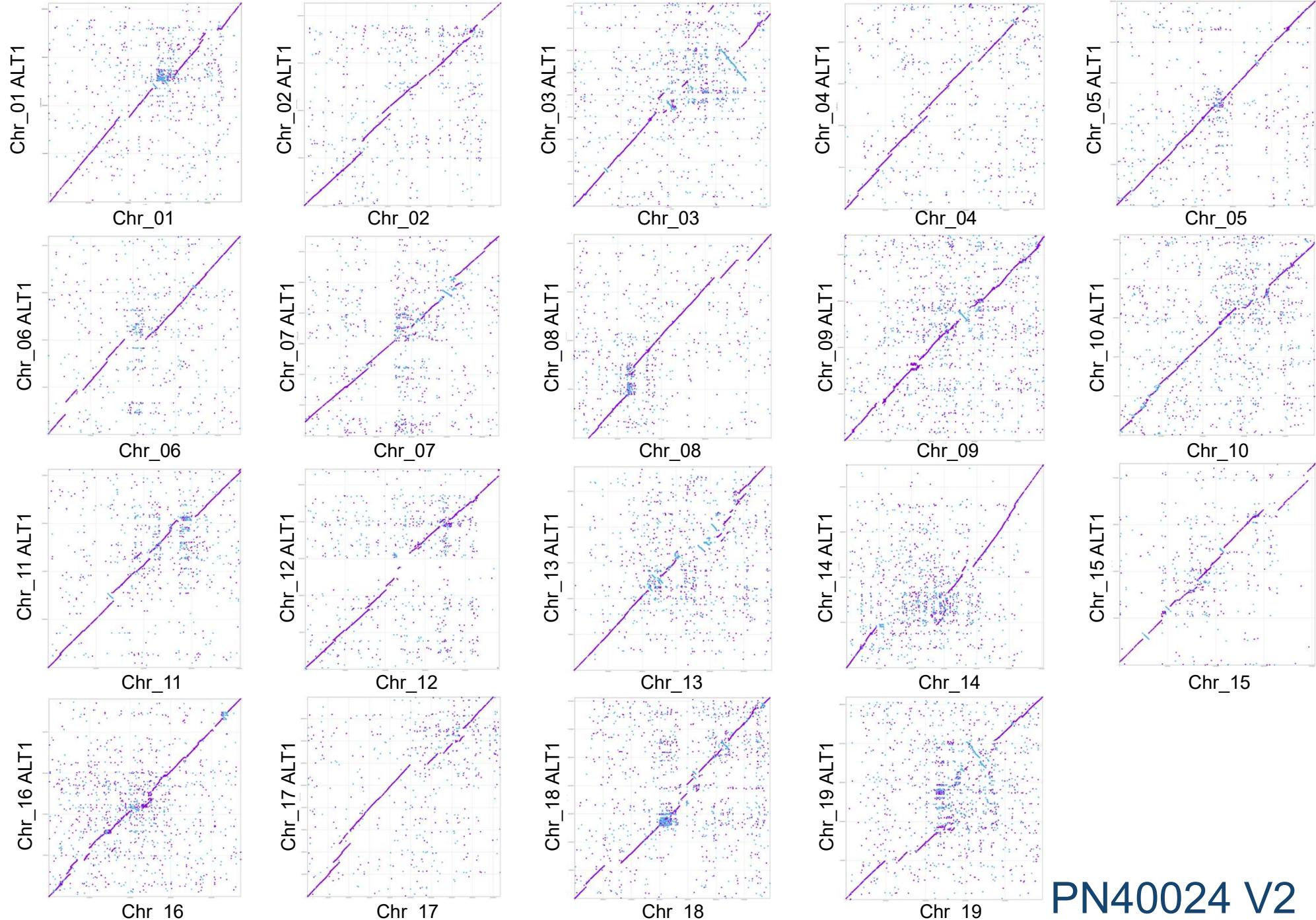
Assembly	Size (Mb)	N. seqs	N50 (L50)	N90 (L90)
Contigs	591	718	2.17 Mb (72)	0.42 Mb (300)
Scaffolds (i)	592	330	9.4 Mb (21)	0.96 Mb (87)
Scaffolds (ii)	592	246	11.3 Mb (19)	1.7 Mb (66)
Scaffolds (iv)	559	182	11.8 Mb (18)	2.8 Mb (52)
Hyperscaffolds ALT1	443	56	16.5 Mb (11)	6.9 Mb (26)
Hyperscaffolds ALT2	330	33	14 Mb (10)	6.1 Mb (23)

Optical maps (phasing and scaffolding)



Alignment to PN40024 and pseudomolecules

Cabernet Sauvignon



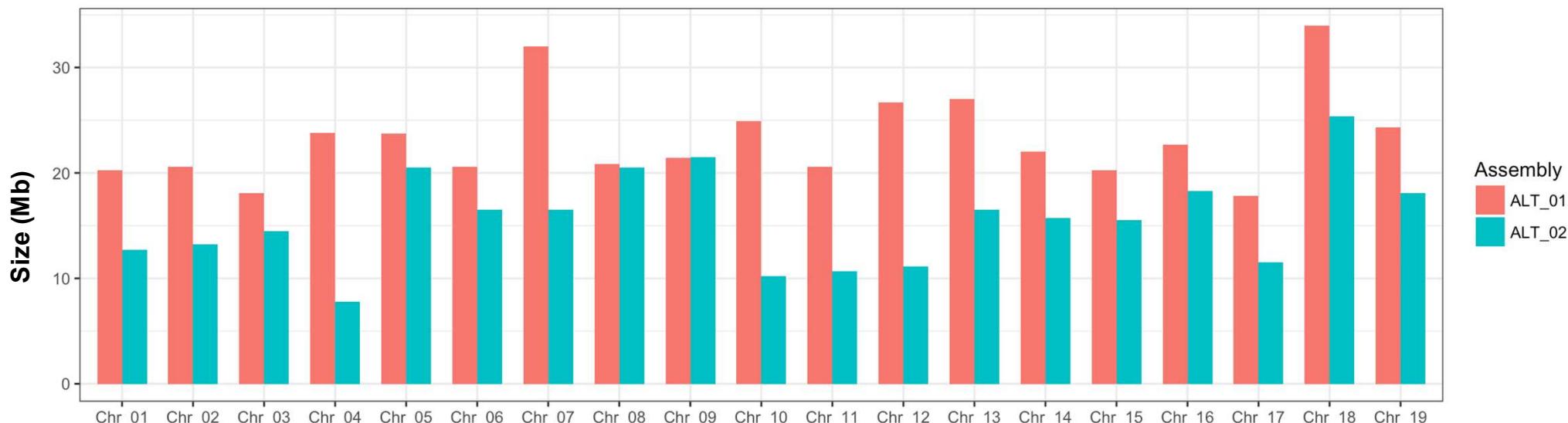
PN40024 V2

68% of the whole genome is phased in two haplotypes

Assembly	Size (Mb)	N. chr	N. contigs	Gaps (Mb)
ALT_1	455.7	19	525	14.2 (3.1%)
ALT_2	310.0	19	422	13.5 (4.3%)

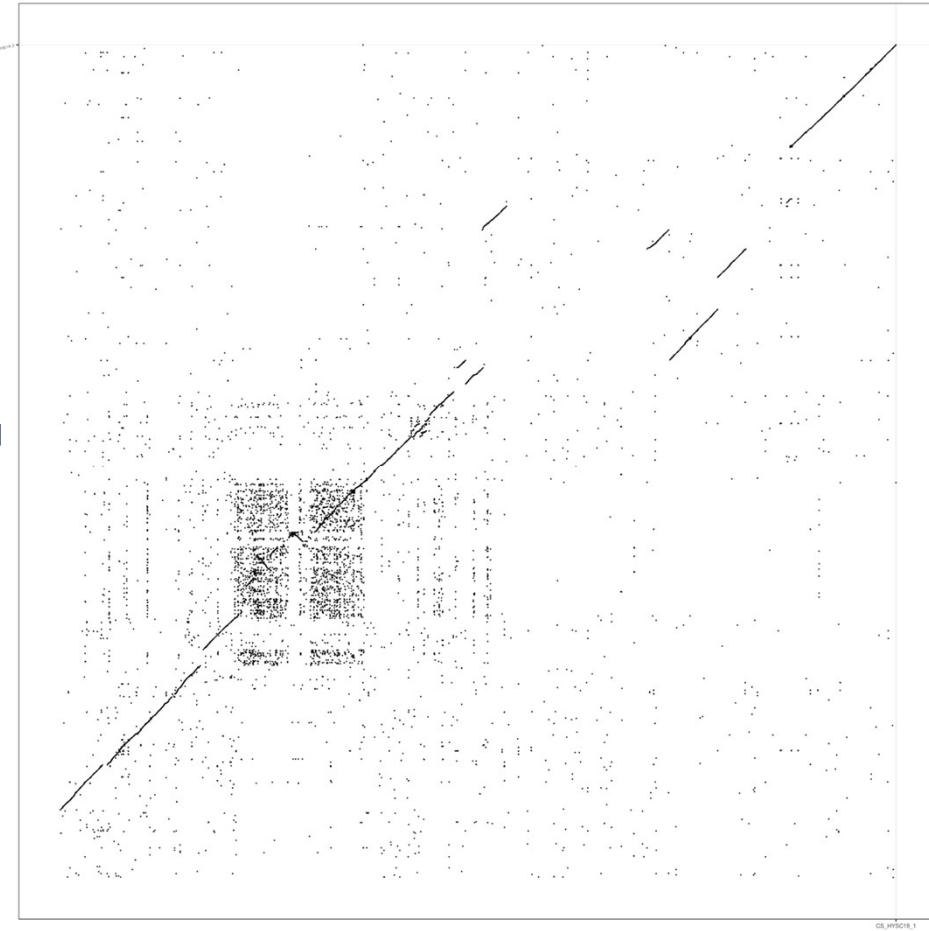
Protein coding sequences

	ALT1	ALT2
Number of CDS	29,294	16,806
Mean CDS length (kb)	1.2	1.2
Mean number of exons / CDS	5	5



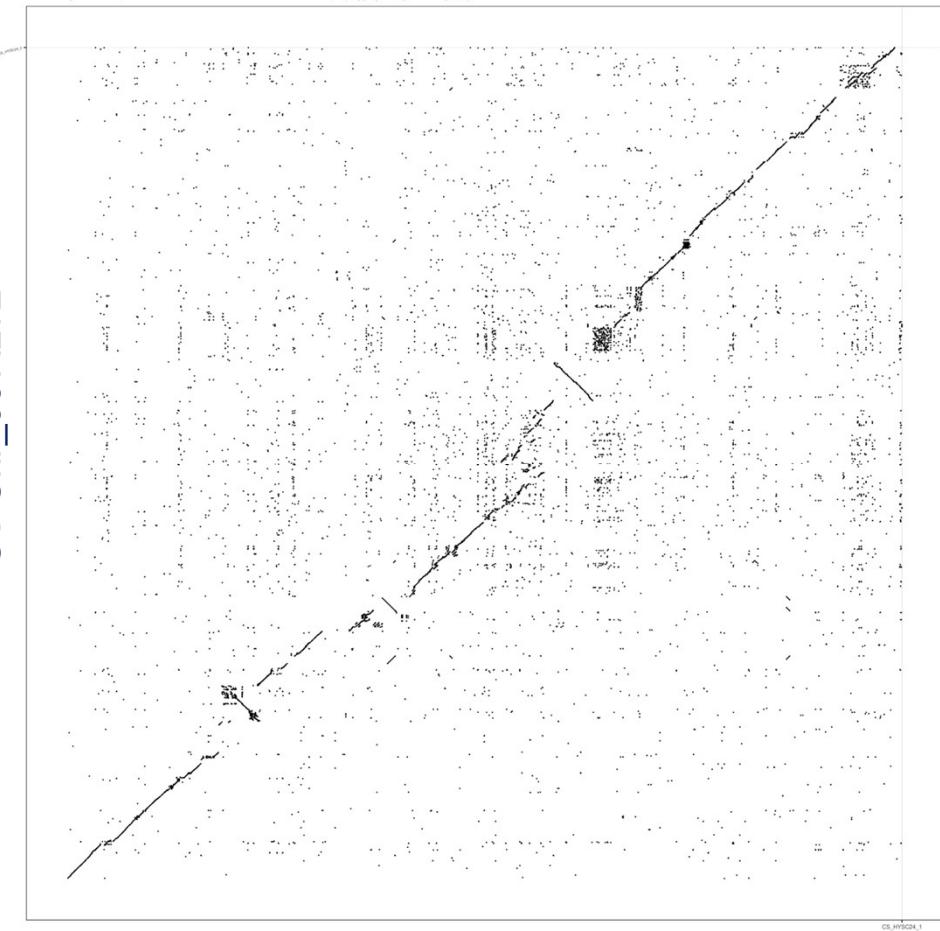
Structural comparison between homologous chromosomes

Chromosome 08



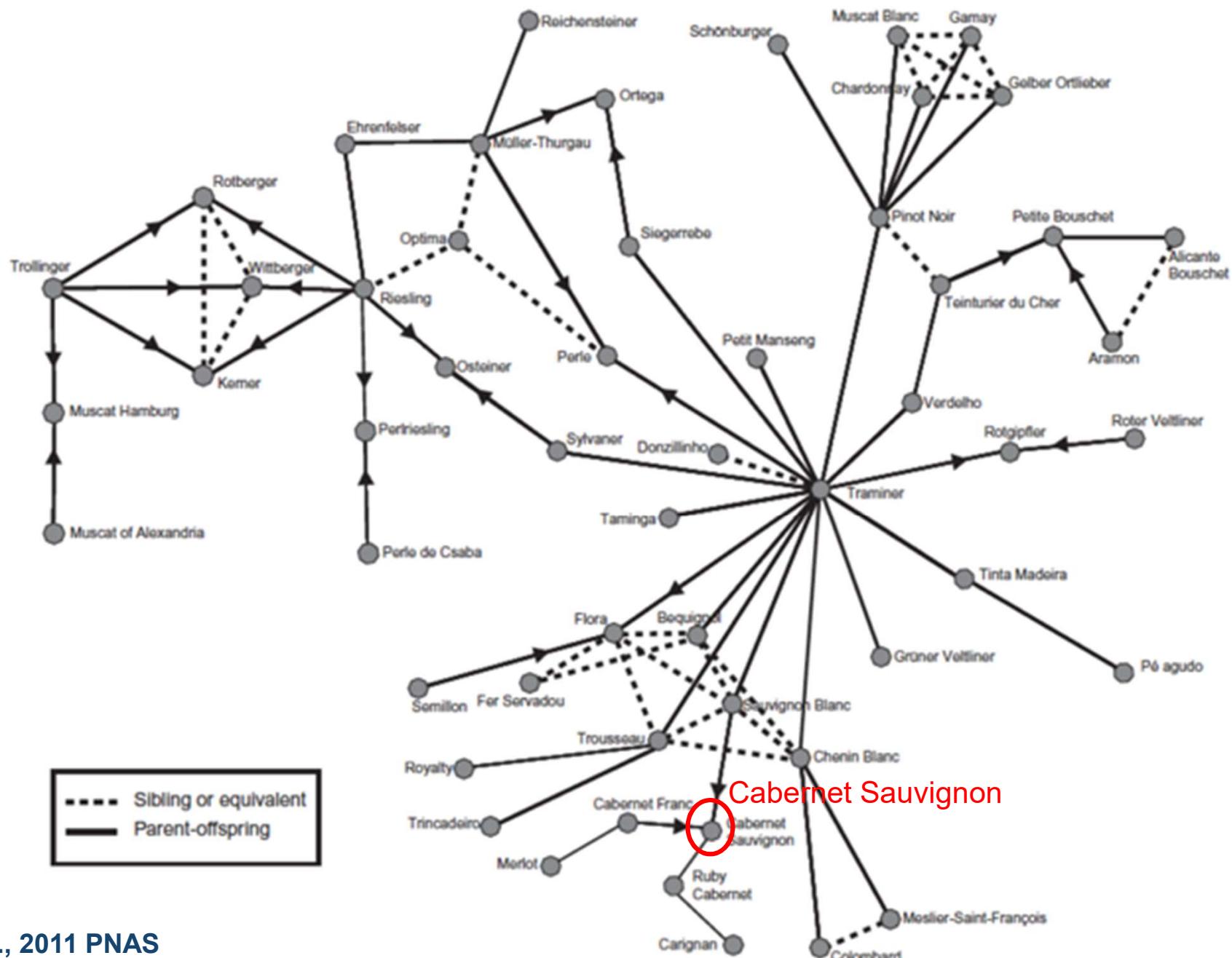
CS Chr_08 ALT1

Chromosome 09



CS Chr_09 ALT1

3. Where are we and where are we going?



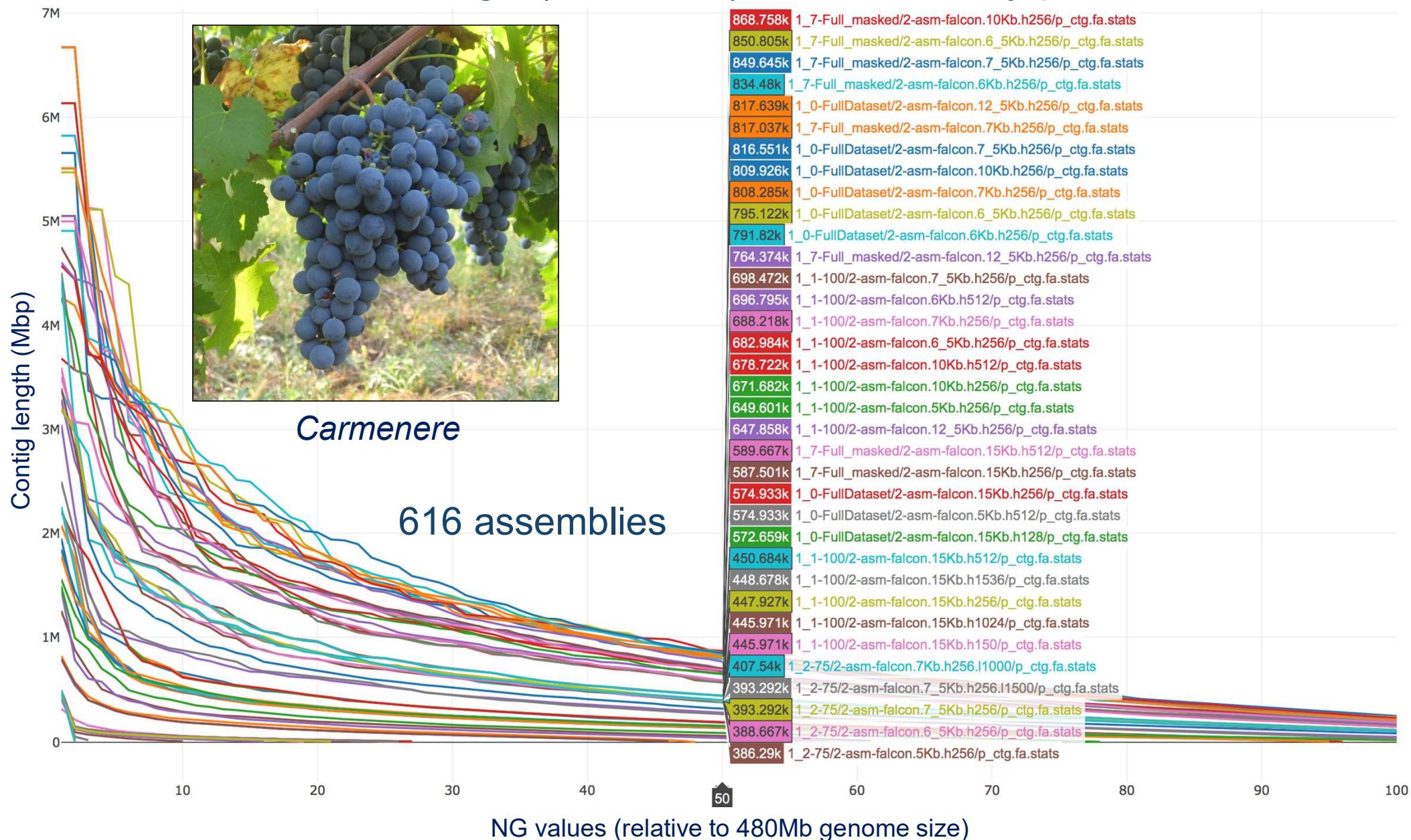
Myles et al., 2011 PNAS

Fig. 3. Network of first-degree relationships among common grape cultivars. Solid vertices represent likely parent-offspring relationships. Dotted vertices represent sibling relationships or equivalent. Arrows point from parents to offspring for inferred trios (details are provided in Materials and Methods).

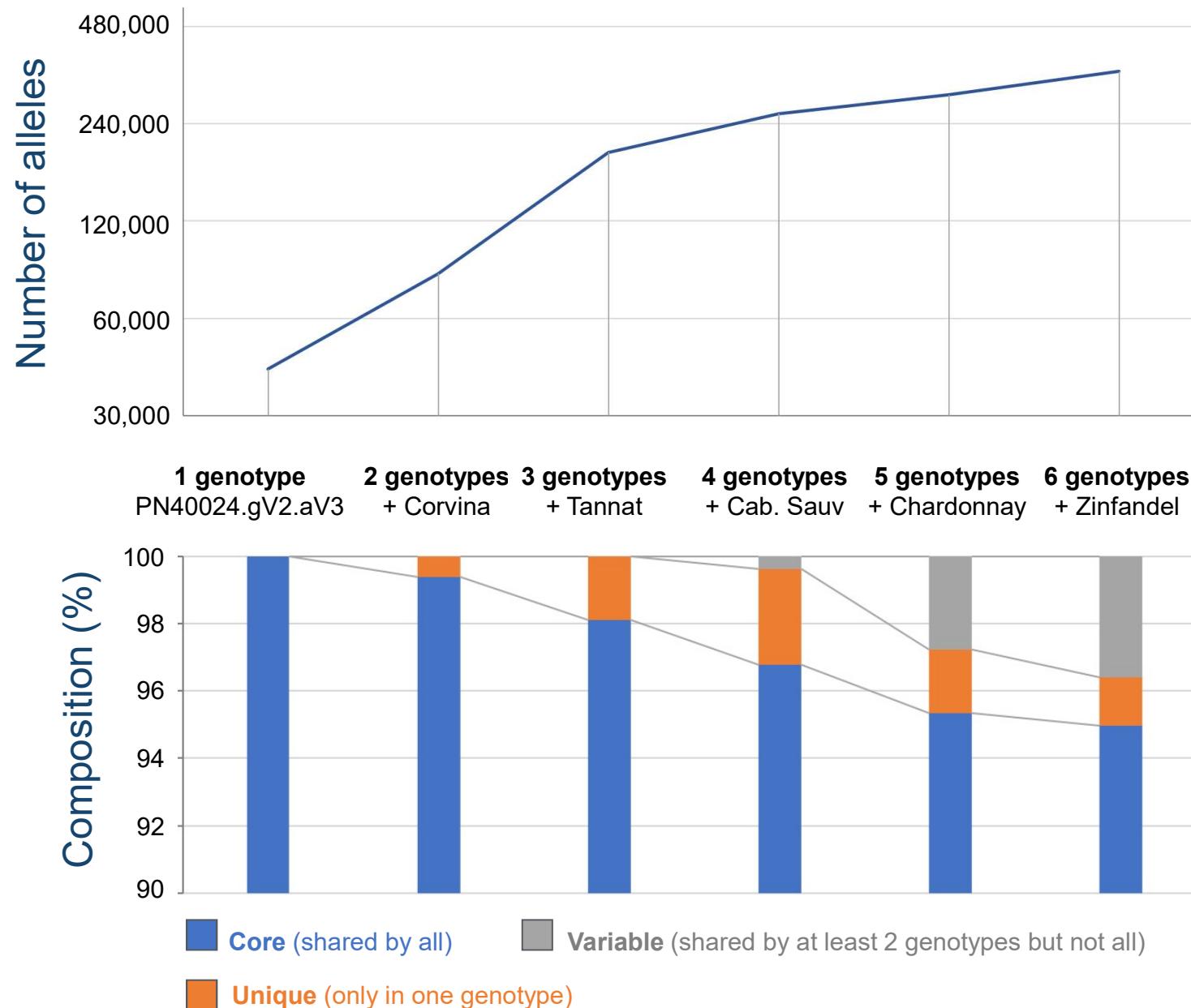


Optimization of SMRT sequencing and FALCON assembly

Different coverage (5x - 115x) and assembly parameters



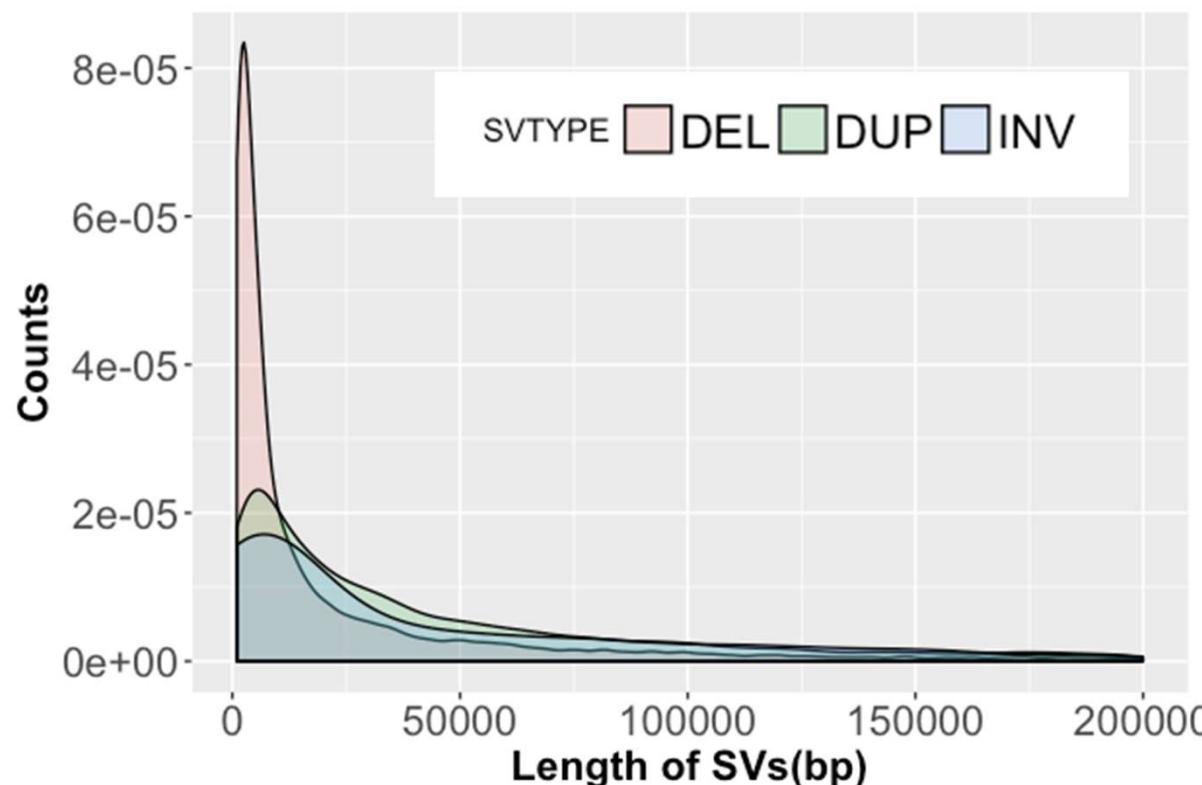
Expanding the gene space



Genomic structural variation

	SNPs + indels*	SVs*
N. variants in <i>V. vinifera</i> spp. <i>sylvestris</i>	5.36 M	0.21 M
N. variants in <i>V. vinifera</i> spp. <i>vinifera</i>	4.91 M	0.19 M
Total size	7.44 Mbp	14.28 Mbp

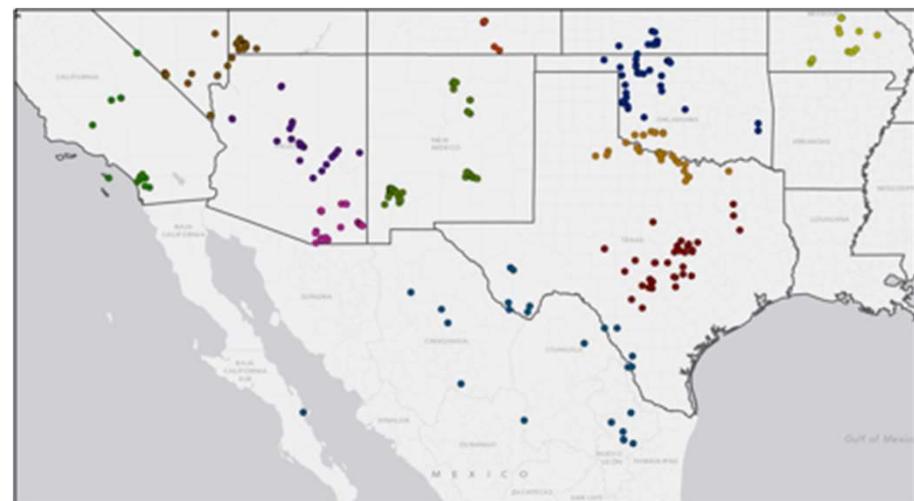
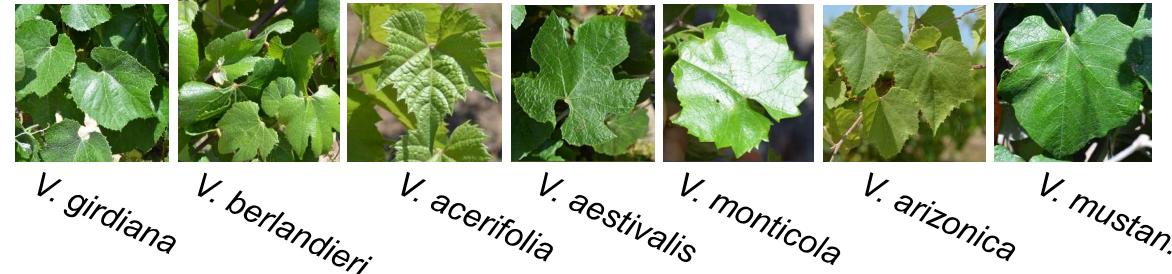
* relative to Chardonnay





What's Next? North American Vitis

Southwest Vitis



A. Pierce's Disease



B. Salt Tolerance



RESEARCH-PGR
#1741627

Ren2
V. cinerea x V. rupestris

Ren5

Run1

Run2
M. rotundifolia

Ren3
american sp.
('Regent')



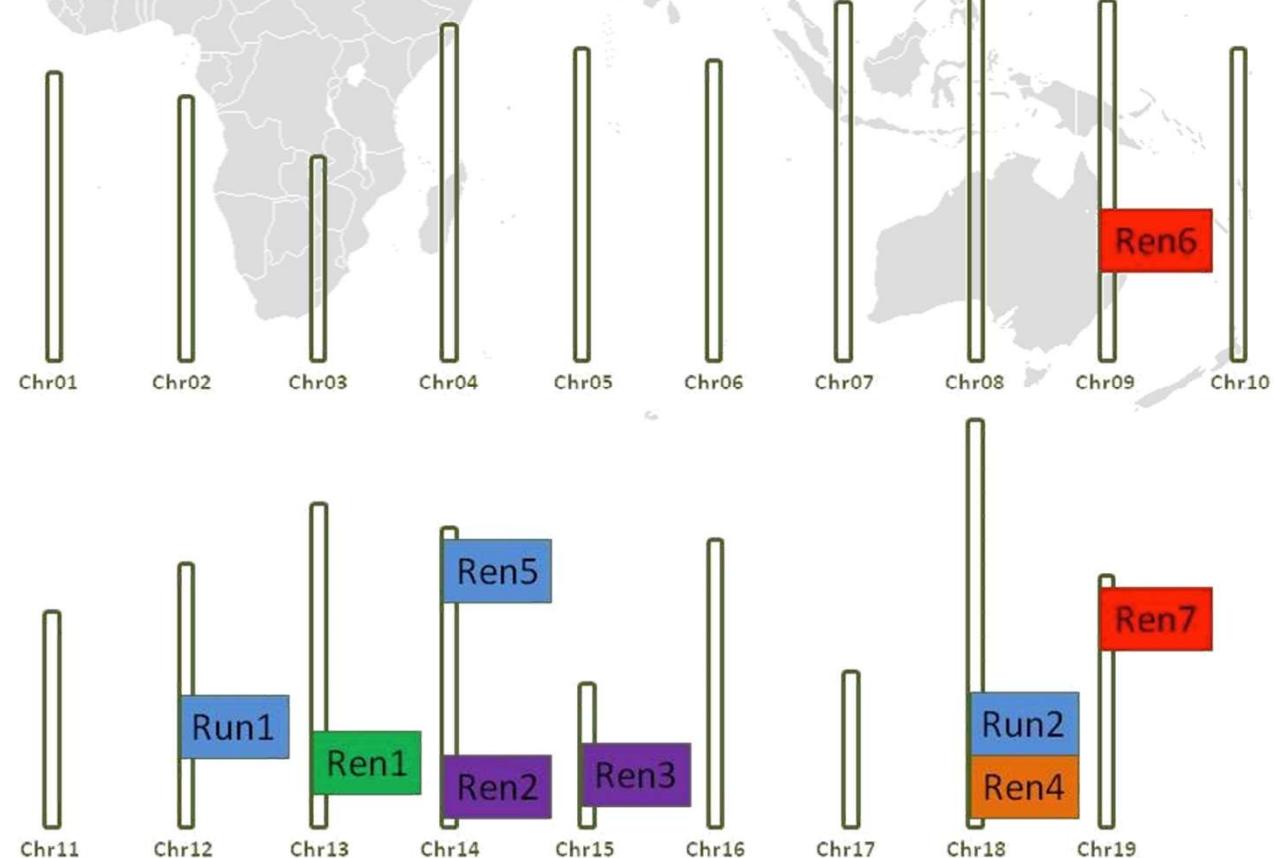
Ren1
V. sylvestris/vinifera

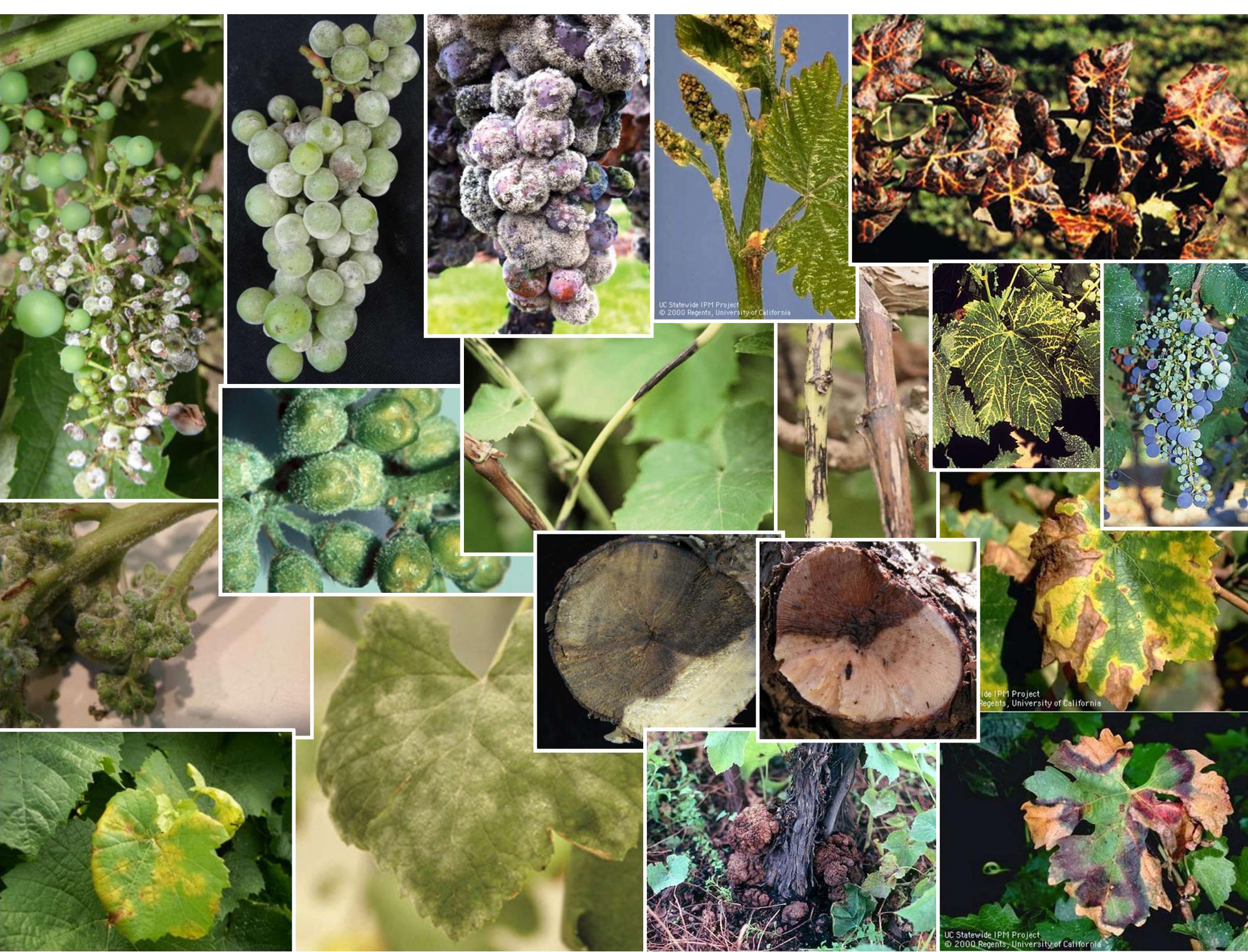
Ren4
V. romeneti

Ren6

Ren7
V. piazeskii

Grape Genome





Species	Disease	Isolates sequenced	Assembly size (Mb)	Number of contigs	N50 (Kb)	Number of genes	Citation
<i>Eutypa lata</i>	Eutypa dieback	11	54.5	10	6,542.31	15,313	Blanco-Ulate et al., 2013a Morales-Cruz et al., 2015
<i>Neofusicoccum parvum</i>	Botryosphaeria dieback	16	43.7	27	2,555.42	13,124	Blanco-Ulate et al., 2013b Morales-Cruz et al., 2015 Massonnet et al., 2016
<i>Phaeoacremonium minimum</i>	Esca complex	5	47.3	24	5,520.70	14,790	Blanco-Ulate et al., 2013c Morales-Cruz et al., 2015 Massonnet et al., 2018
<i>Phaeomoniella chlamydospora</i>	Esca complex	2	27.5	702	178.60	6,986	Morales-Cruz et al., 2015
<i>Diplodia seriata</i>	Botryosphaeria dieback	1	37.1	695	304.20	9,398	Morales-Cruz et al., 2015
<i>Diaporthe ampelina</i>	Phomopsis dieback	1	47.4	2,392	132.30	10,801	Morales-Cruz et al., 2015
<i>Erysiphe necator</i>	Powdery mildew	5	52.5	5,936	21.4	6,533	Jones et al., 2014

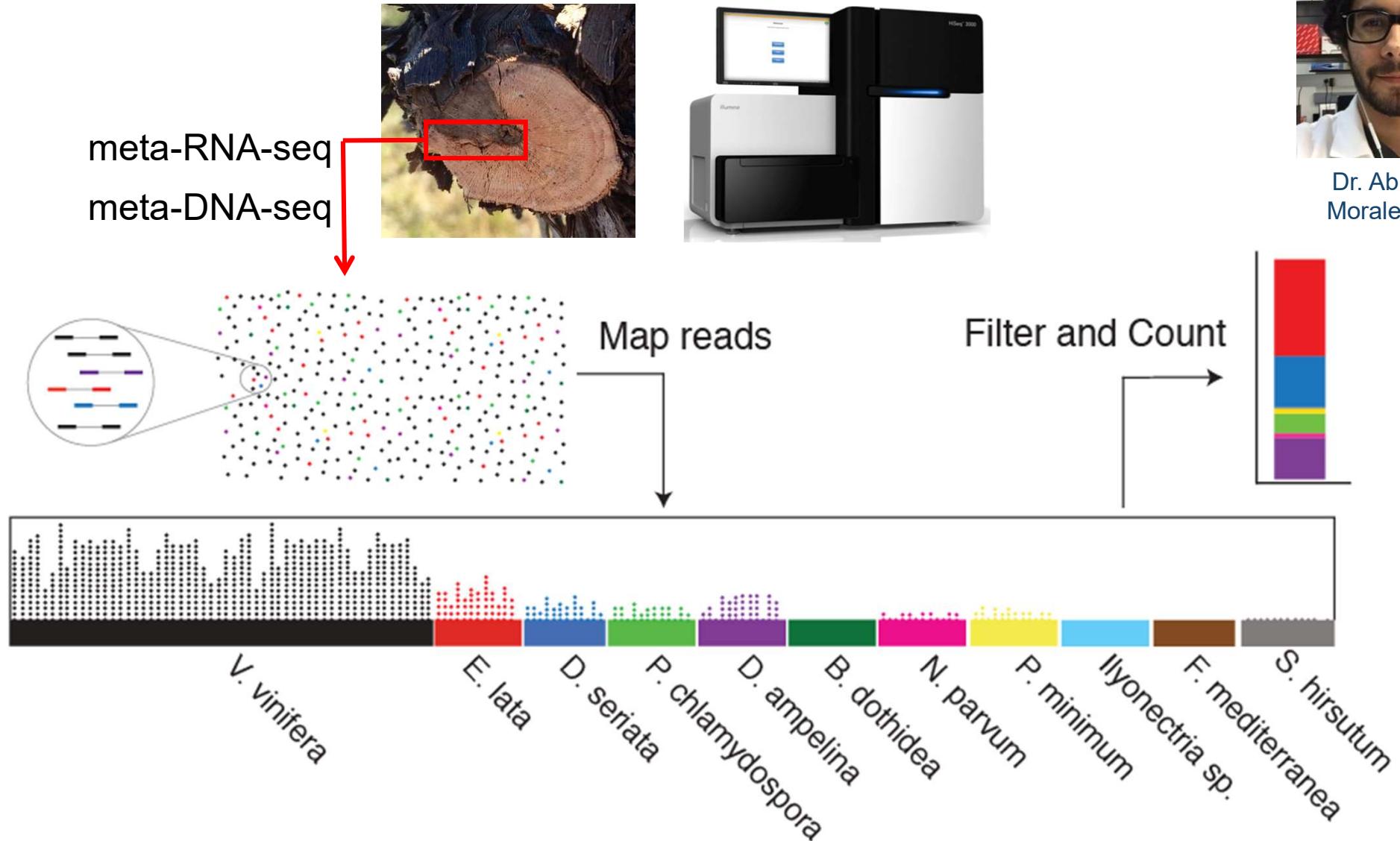




The vineyard metagenome



Dr. Abraham
Morales-Cruz



Multi-species reference for RNA-seq read mapping

Sampling genetic diversity

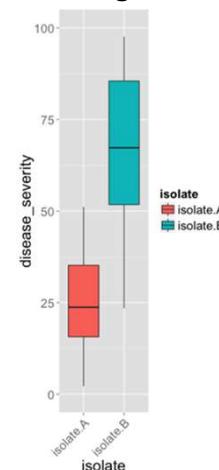
Multiple hosts



Multiple geographical locations



Different virulence strengths



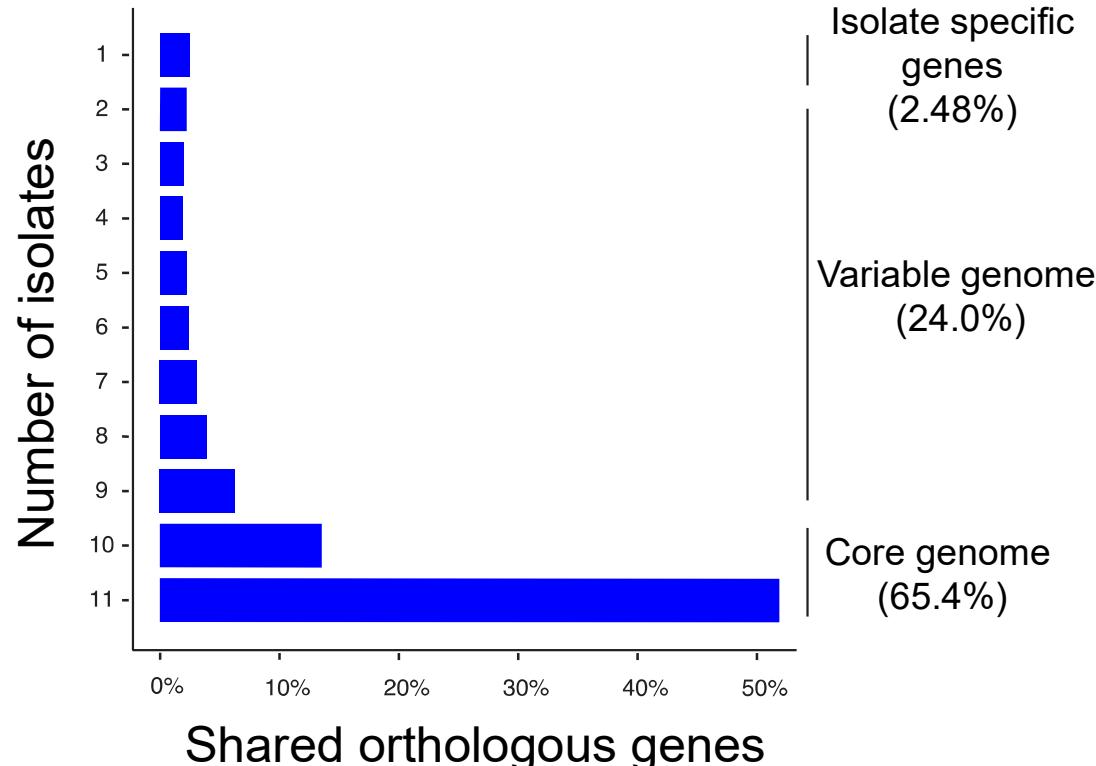
AA9 family gene member

- ES302.7.g7941 AA9
- EL1.72.g13343 AA9
- EN210.1.g187 AA9
- ES015.12.g11084 AA9
- ES213.115.g4359 AA9
- EN310.79.g1171 AA9
- EN021.3.g685 AA9 ●
- ES201.17.g11271 AA9
- EN209.16.g11266 AA9
- EN006.4.g7671 AA9
- EI EN209 br gene g3749.t1
- ES012.72.g2287 AA9
- EN021.216.g8882 AA9 ●

- CAZyme with Auxiliary Activity
- Copper-dependent lytic polysaccharide monooxygenases (LPMOs)
- Cleaves cellulose chains with oxidation of various carbons (C-1, C-4 and C-6)

Host	Host Species	County	Agressivness	Total Assembly Size (Mb)	Number of Scaffolds	N50 (Kb)
Grape	<i>Vitis vinifera</i>	Napa Co.	low	54.52	789	275.15
Grape	<i>Vitis vinifera</i>	Napa Co.	high	54.46	819	284.26
Apricot	<i>Prunus armeniaca</i>	Napa Co.	high	54.27	721	284.26
Apricot	<i>Prunus armeniaca</i>	Napa Co.	low	54.40	675	264.43
Willow	<i>Salix sp</i>	Napa Co.	unknown	51.50	787	290.78
Grape	<i>Vitis vinifera</i>	San Benito Co.	high	54.39	663	282.35
Grape	<i>Vitis vinifera</i>	San Benito Co.	low	54.48	739	264.69
Apricot	<i>Prunus armeniaca</i>	San Benito Co.	low	54.58	905	245.54
Apricot	<i>Prunus armeniaca</i>	San Benito Co.	high	54.67	888	245.95
Willow	<i>Salix sp</i>	San Benito Co.	unknown	54.59	778	239.54

Gene cluster conservation across isolates



The take home message

We are going toward one (diploid) genome reference per grape species/cultivar/accession.



What do we still need?

1. Data sharing, organization, and retrieval
2. How do we ensure that results are comparable across experiments?
3. Bioinformatic tools with diploid references
4. Data visualization: from “genome-browsers” to “multiple haplotype-browsers”



VITICULTURE & ENOLOGY
UNIVERSITY OF CALIFORNIA DAVIS



Dr. Mélanie
Massonnet



Dr. Abraham
Morales-Cruz



Dr. Amanda Vondras



Dr. Rosa
Figueroa



Dr. Andrea Minio



Jerry Lin



Shahin S. Ali



Yang He



Jadran Garcia



Daniela Quiroz



Lucero Espinoza



Dingren Yang

<http://cantulab.github.io/>

@cantulab   

Acknowledgments

Collaborators



Andy Walker (UCD)
Brandon Gaut (UCI)
Grant Cramer (UNR)
Kendra Baumgartner (USDA-ANR)
Philippe Rolshausen (USDA-ANR)
Massimo Delledonne (Univ. Verona)
Lance Cadle Davidson (USDA-ANR)
Jason Londo (USDA-ANR)

Funding



E&J Gallo Winery

MARS



CORFO



J. LOHR
VINEYARDS & WINES



CALIFORNIA DEPARTMENT OF
FOOD & AGRICULTURE



Liquid Gold from Napa Valley



College of Agricultural & Environmental Sciences