

# *De-novo assembly of six grapevine cultivars reveals novel gene elements*

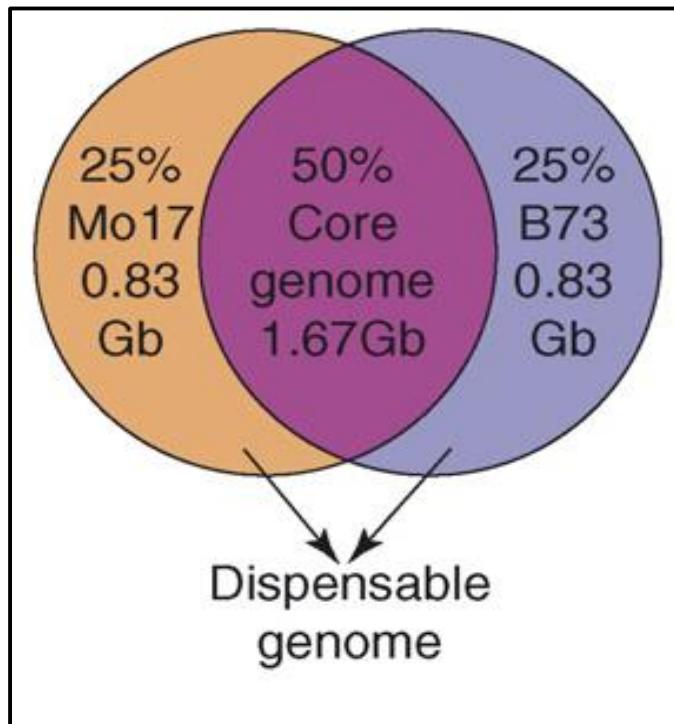
Michele Vidotto

*michele.vidotto@uniud.it*



image: camera di commercio di Bolzano

# The pan-genome concept

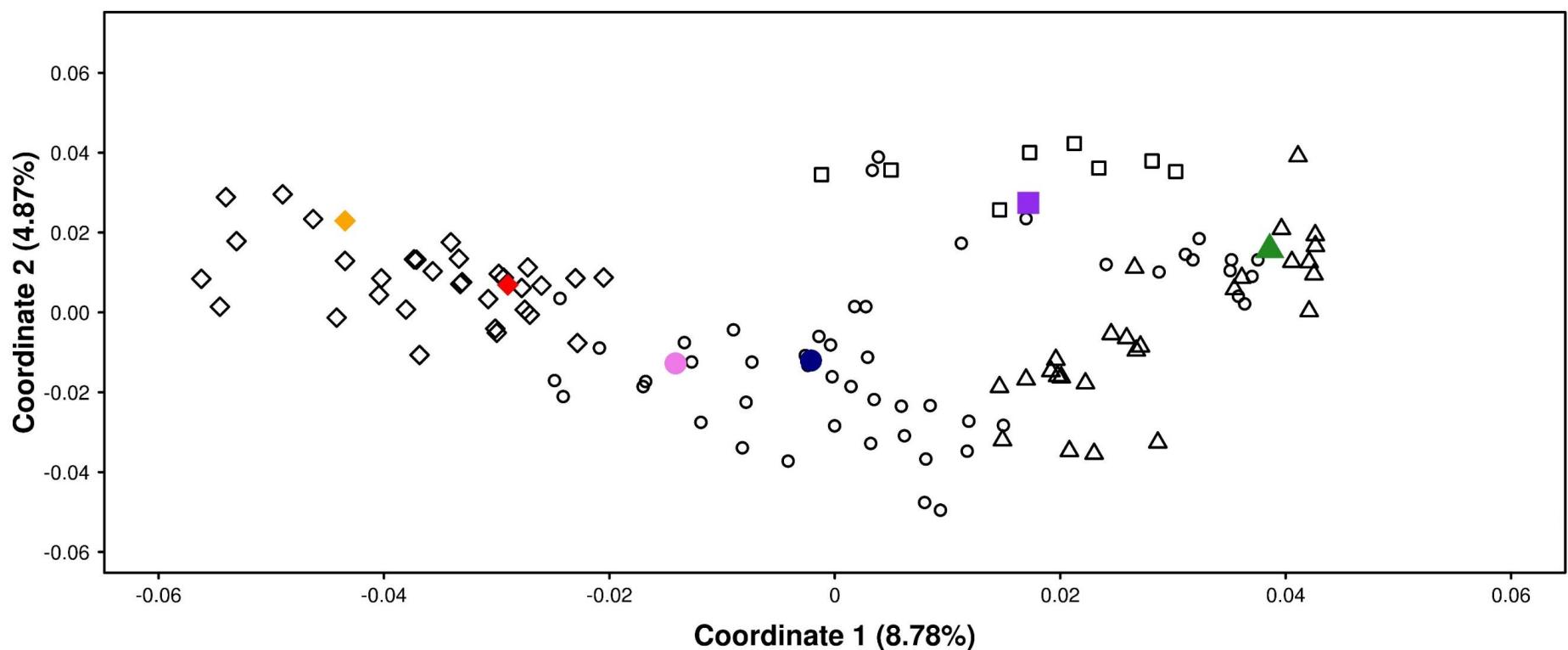


Morgante *et al.* Curr Opin Pl Biol, 2007

- In plants, the dispensable genome is mostly composed by transposable elements (TE)
- A gene-like fraction can be also found
- Characterisation of *Vitis Vinifera* pan-genome genes

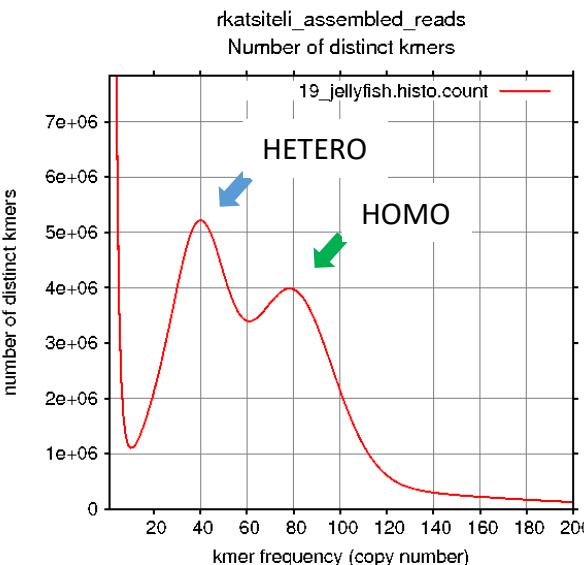
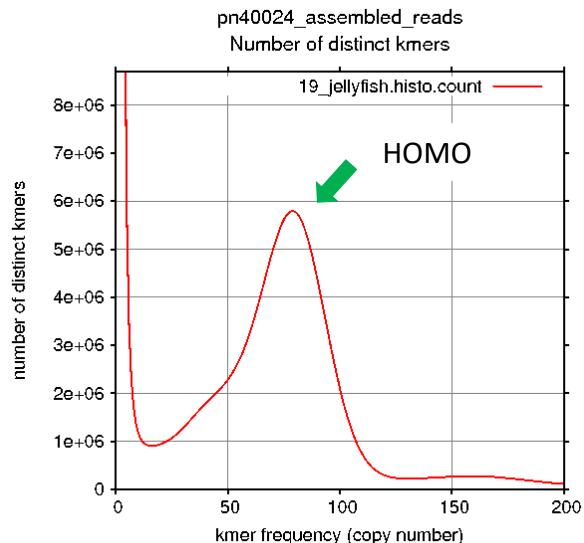
# Choice of samples that maximize genetic diversity

- ◆ *Cabernet franc*
- ◆ *Traminer*
- *Sangiovese*
- *Gouais blanc*
- ▲ *Kishimish vatkana*
- *Rkatsiteli*

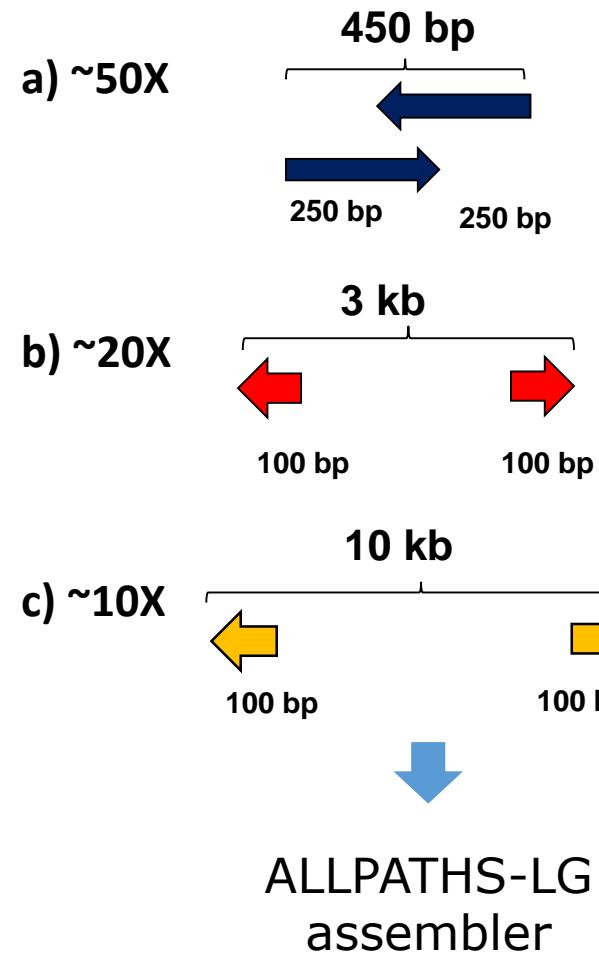


# Facing the heterozygosity problem

- 19 chr
- sanger reference  
(PN40024)
- 487 Mb
- highly heterozygous individuals



- 3 sequencing library



# Good assembly results

	Traminer	Gouais blanc	Cabernet franc
Total bp	678,662,631	743,730,069	711,707,043
Sequences (#)	8,676	9,564	12,485
N50 (#)	484	524	1,178
L50 (bp)	413,025	408,071	178,391

	Rkatsiteli	Kishmish vatkana	Sangiovese	Illumina reference
Total bp	753,744,202	610,410,426	691,235,812	521,491,794
Sequences (#)	10,089	8,546	15,667	6,016
N50 (#)	612	457	627	553
L50 (bp)	352,572	364,852	301,183	264,099

# Gene Prediction and structural annotation

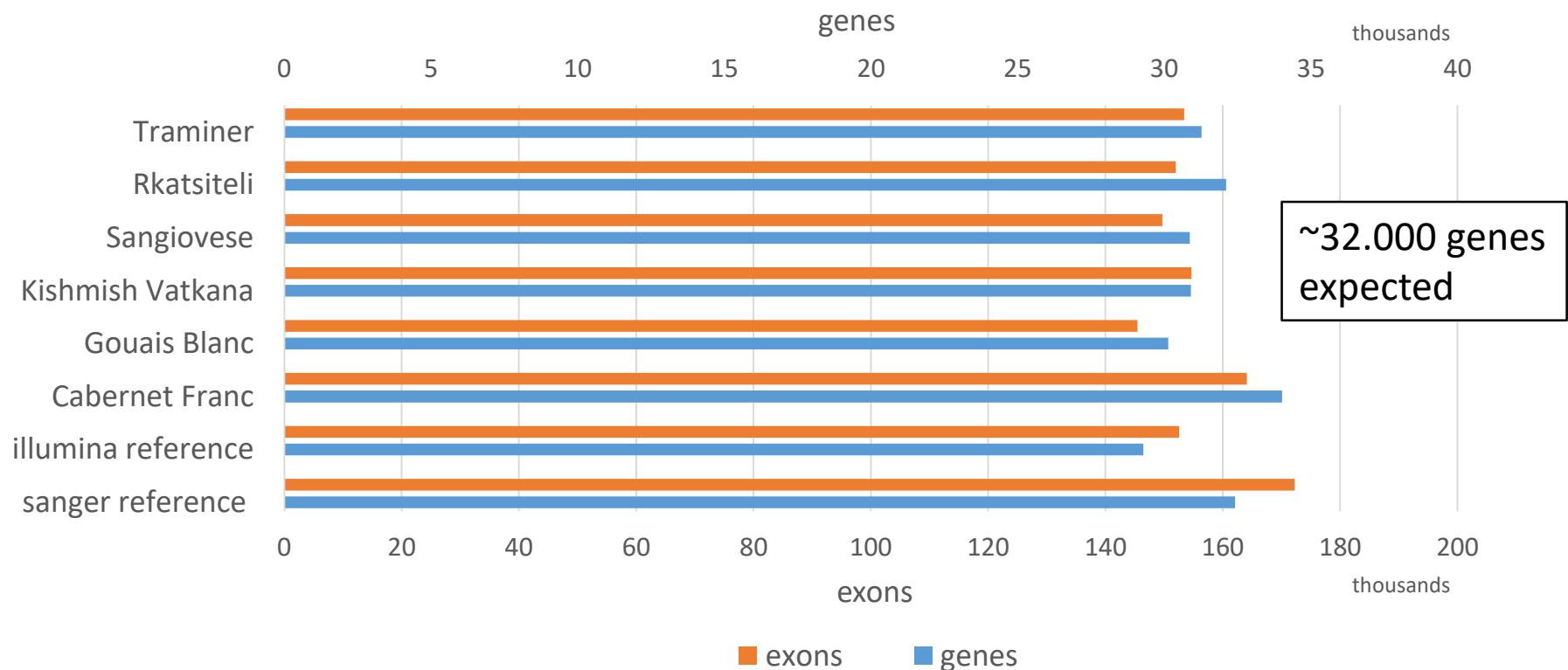
Evidences:

- Sample specific RNA-Seq from leaves, tendrils and berries
- Collection of *V. vinifera* EST & protein (NCBI)
- Reference transcripts (v2.1)

- *De-novo*
- Evidence based



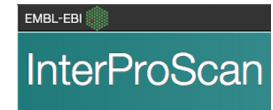
Structural annotation statistics of protein coding genes



# Functional annotation and validation

## Functional annotation:

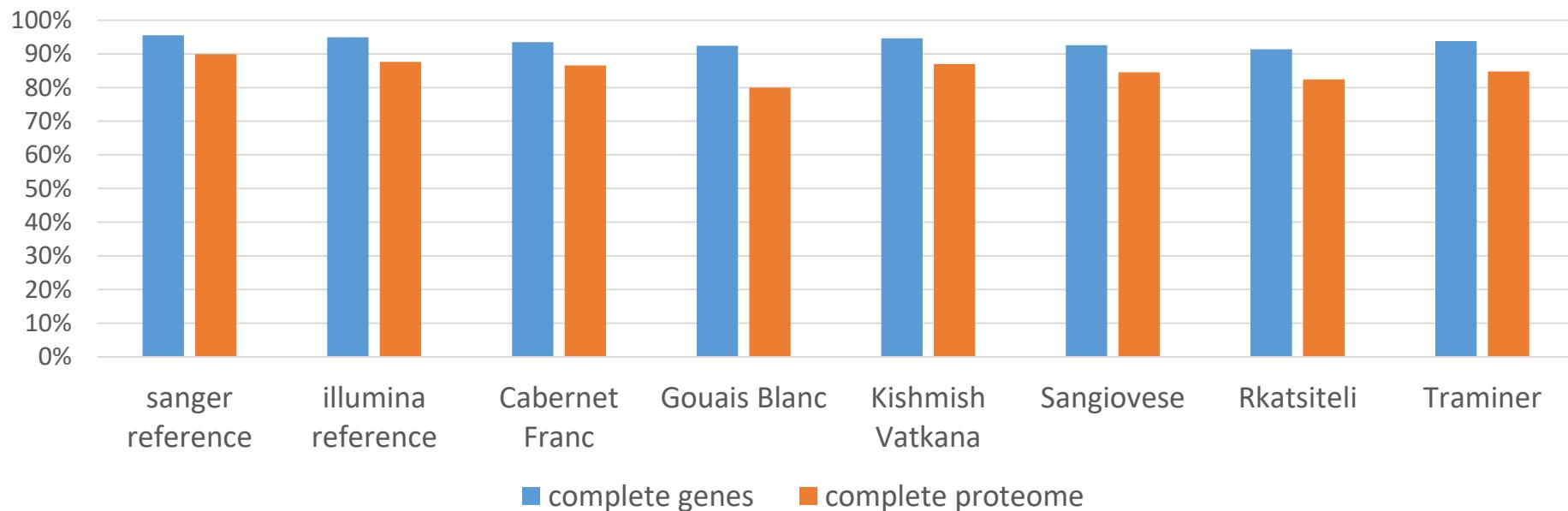
- Search for **protein signature of conserved domains**
- Corresponding **GO terms** mapping and annotation



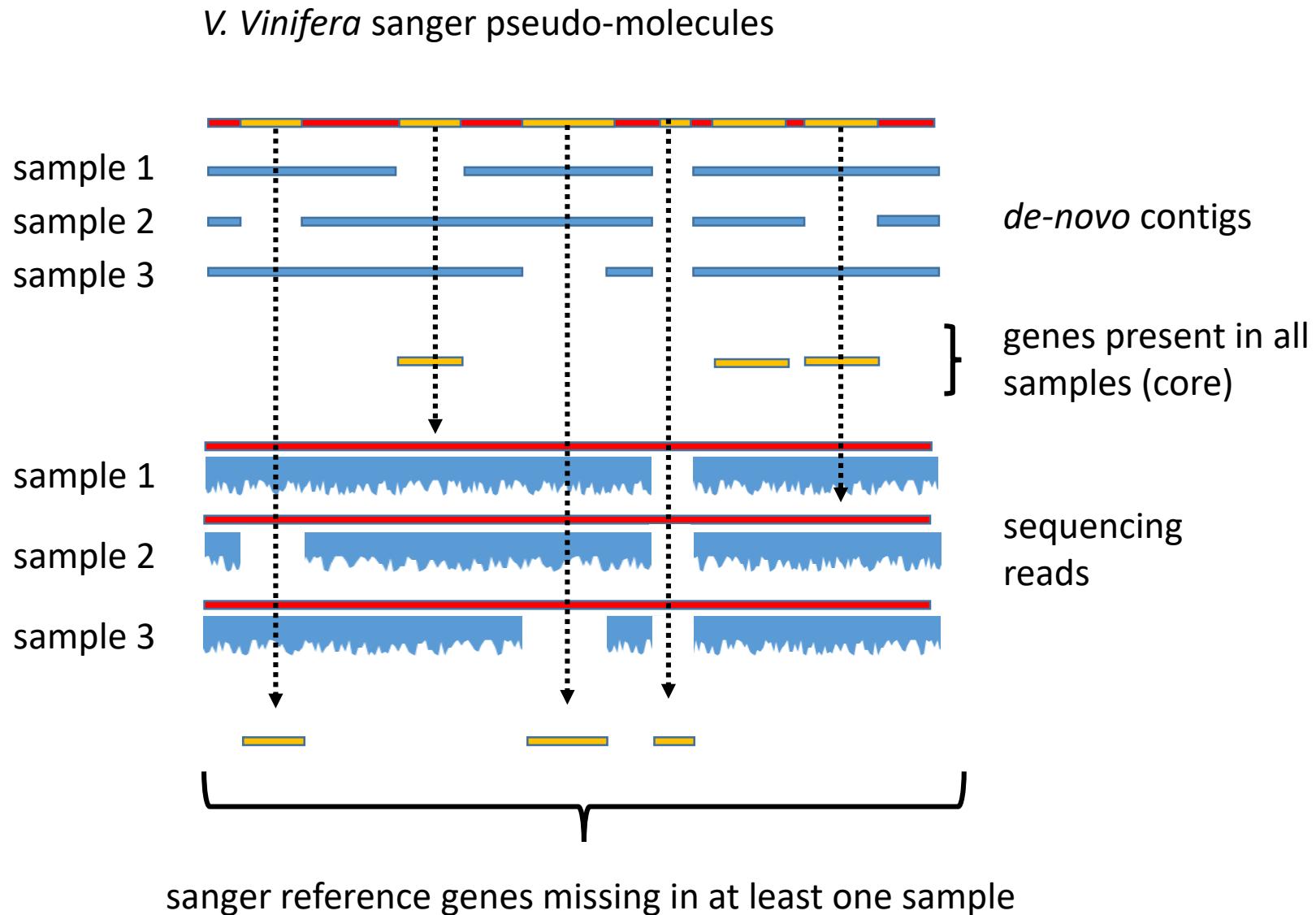
## Validation:

- Check for the presence of plant-specific **Universal Single-Copy Orthologs**

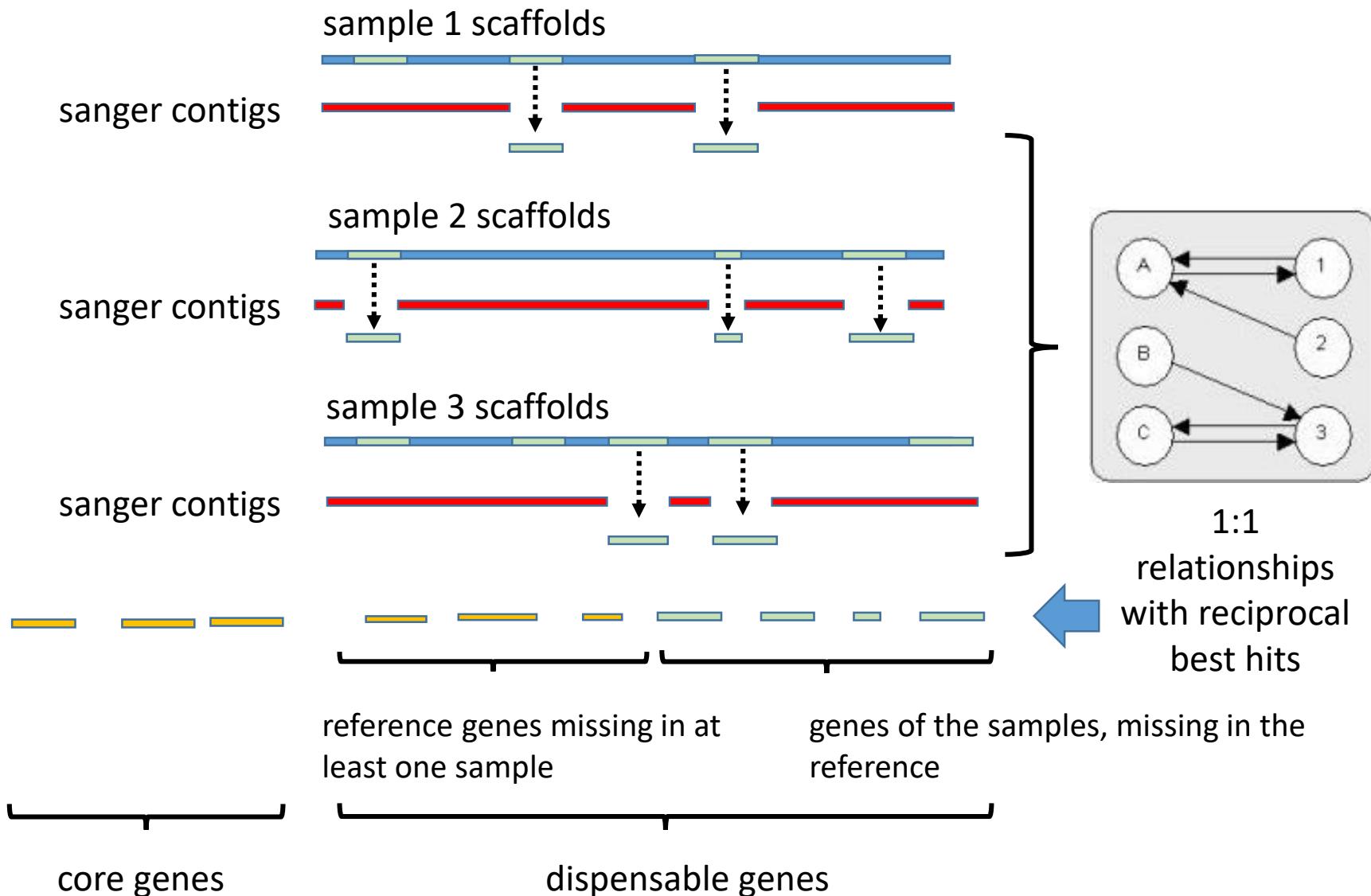
Universal Single-Copy Orthologs found



# Strategy for finding shared and private genes - 1

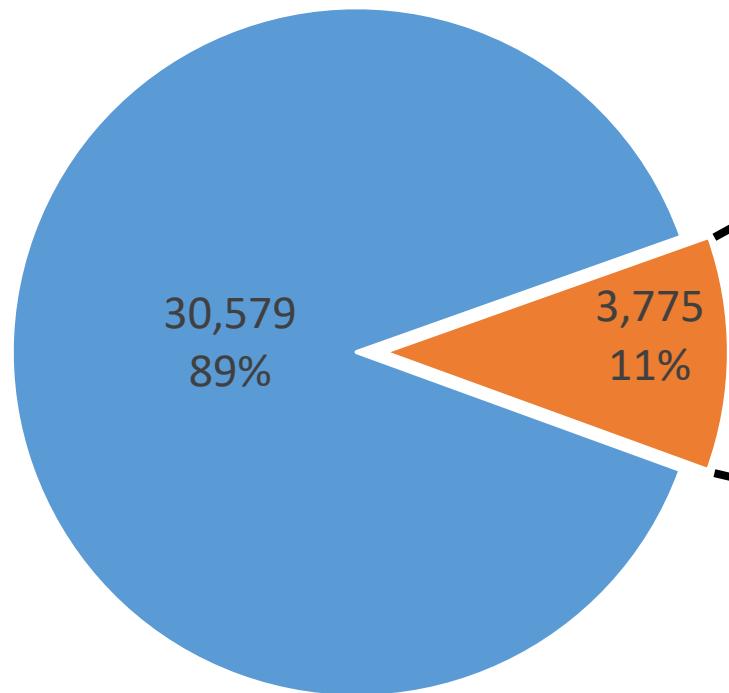


# Strategy for finding shared and private genes - 2

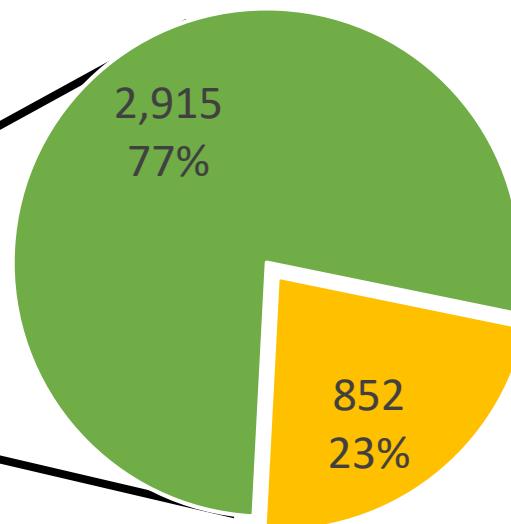


# *Vitis Vinifera* pan-genome gene sets composition

*Vitis vinifera* pan-genome genes



Disposable genes



Traminer
Sangiovese
Rkatsiteli
Kishmish Vatkana
Gouais Blanc
Cabernet Franc

98
118
121
157
117
249

- missing in at least one sample
- unique within samples

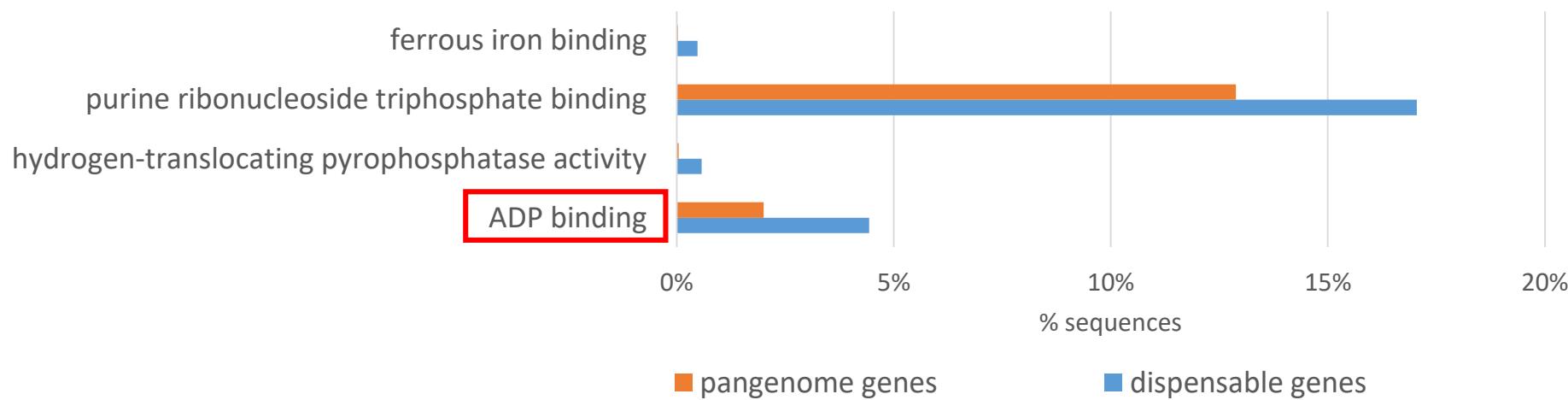
■ core

■ dispensable

# Resistance genes enrichment in the dispensable component

- Functional enrichment based on GO terms
- Fischer's exact test with FDR-corrected p-value (> 0.05)

Differential distribution  
of molecular function GO-terms



Sequences producing significant alignments:

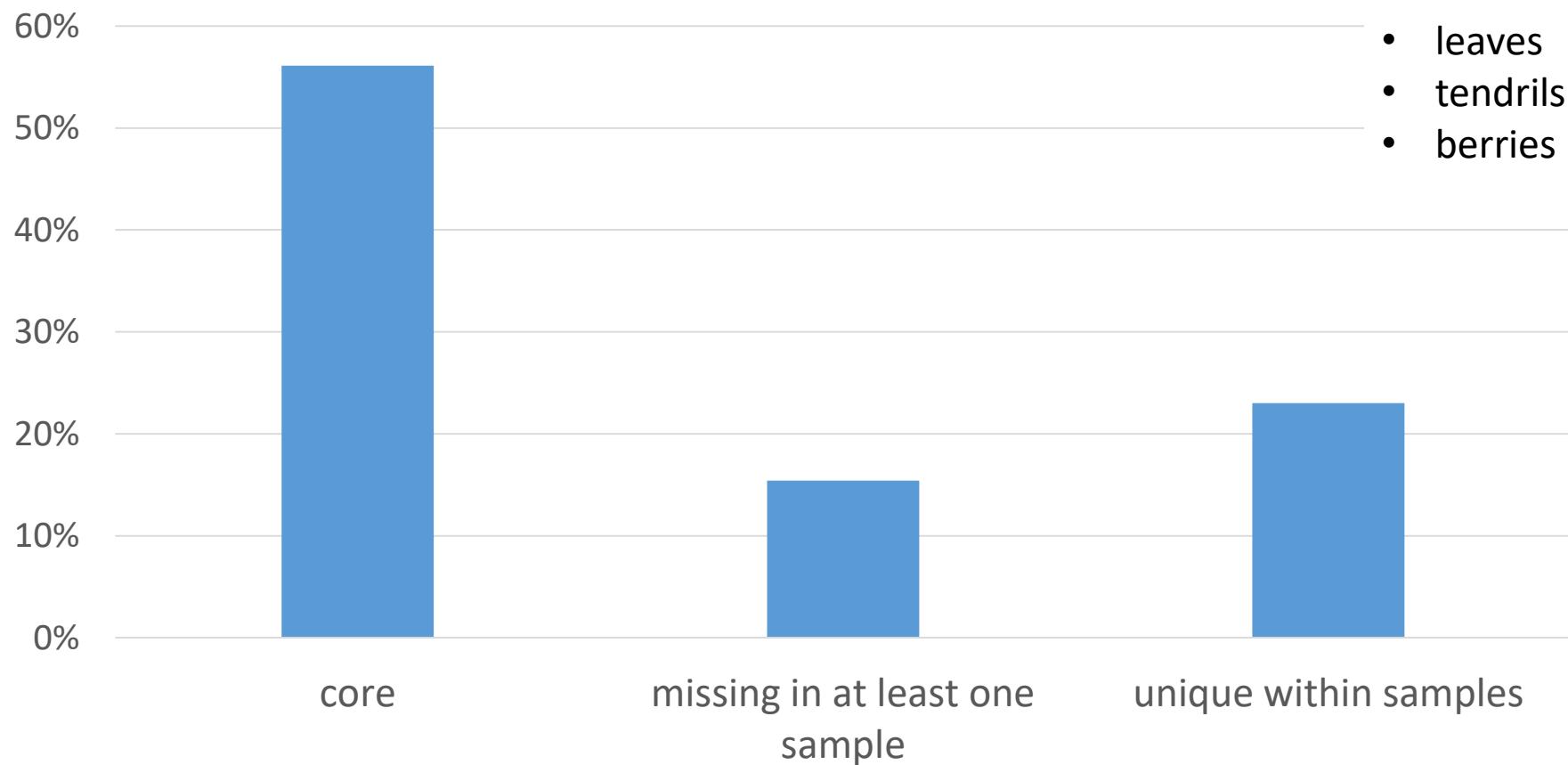
Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Vitis vinifera, whole genome shotgun sequence, contig VV78X124158.82, clone ENTAV 115	4826	5904	97%	0.0	98%	<a href="#">AM487700.1</a>
<input type="checkbox"/>	PREDICTED: Vitis vinifera putative disease resistance protein At1g50180 (LOC100855277), transcript variant X2, mRNA	4540	5462	97%	0.0	97%	<a href="#">XM_019225695.1</a>
<input type="checkbox"/>	Vitis vinifera contig VV78X133772.3, whole genome shotgun sequence	4523	5447	97%	0.0	96%	<a href="#">AM460999.2</a>
<input type="checkbox"/>	PREDICTED: Vitis vinifera putative disease resistance protein At1g50180 (LOC100855277), transcript variant X1, mRNA	4444	5366	95%	0.0	96%	<a href="#">XM_010663478.2</a>
<input type="checkbox"/>	PREDICTED: Vitis vinifera disease resistance protein RPH8A (LOC100258544), mRNA	2837	4973	79%	0.0	100%	<a href="#">XM_019225570.1</a>
<input type="checkbox"/>	Vitis vinifera contig VV78X266359.4, whole genome shotgun sequence	1561	3574	67%	0.0	93%	<a href="#">AM428237.2</a>

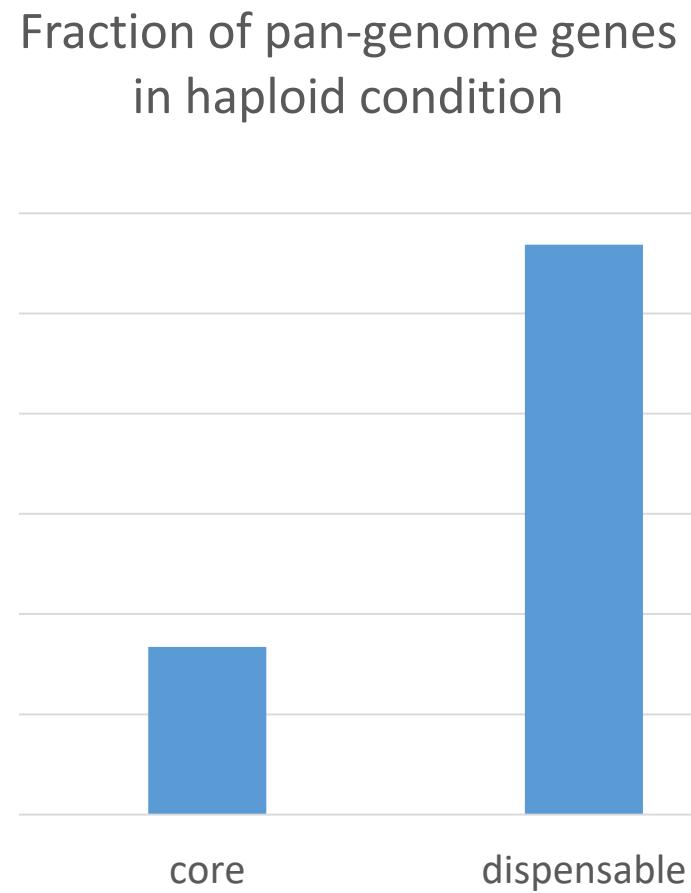
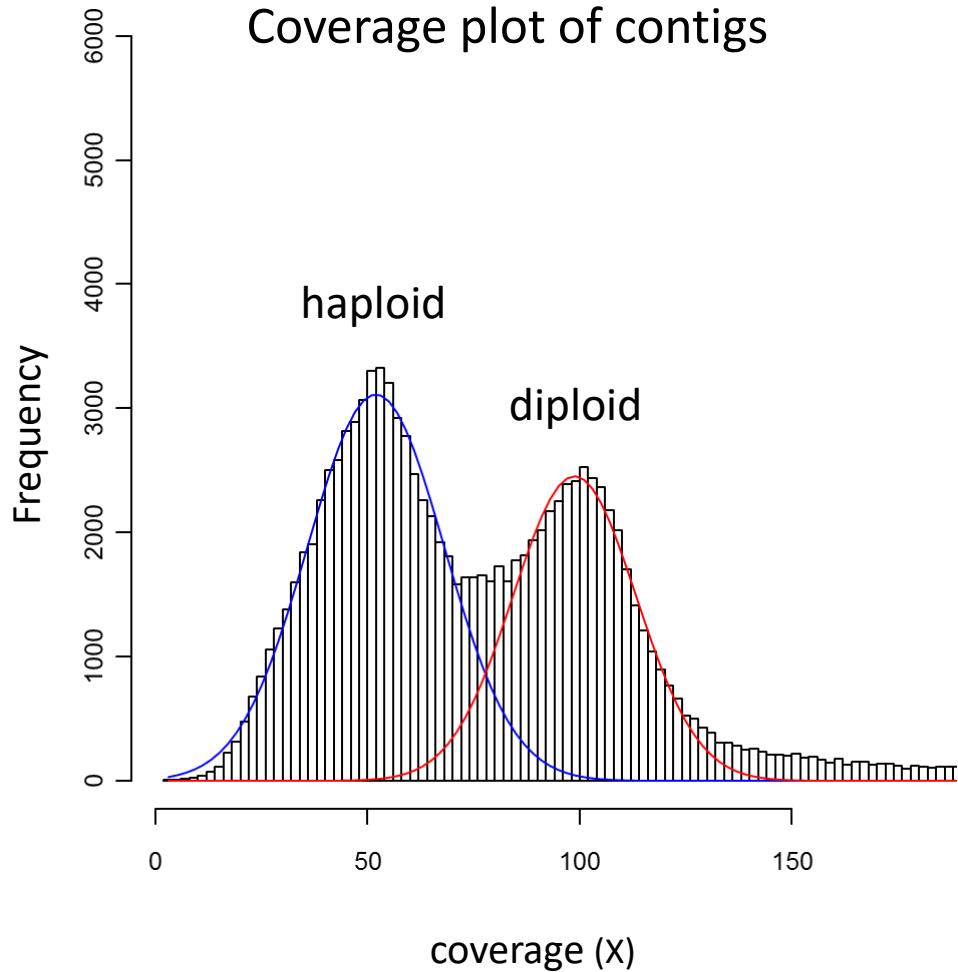
# Expression statistics of the different gene sets

Fraction of pan-genome genes expressed in at least one sample and tissue

(fpkm >=1)



# Fraction of pan-genome genes in single copy



# Conclusion

- Besides SNP and TE induced structural variation, **gene PAV** is another variation form playing an important role in plants.
- With modern assembly algorithms it's possible to reconstruct **high heterozygous genomes** from short reads.
- Variety-specific *de-novo* genome assembly and annotation allow the investigation of a **previously uncharacterized** gene space.
- *De-novo* assembly can also be a proxy to detect the fraction of genome in **haploid** condition.

# Acknowledgments

**UNIVERSITÀ  
DEGLI STUDI  
DI UDINE**



**IGA**   
ISTITUTO DI GENOMICA APPLICATA

Davide Scaglione  
Gabriele Magris  
Sara Pinosio  
Nicoletta Felice  
Giusi Zaina  
Fabio Marroni  
Gabriele Di Gaspero  
Michele Morgante



European Research Council

Established by the European Commission

**Novabreed**  
**ERC-2011-ADG**

