Durum Wheat Pan-Transcriptome

Danara Ormanbekova, Marco Maccaferri , Sven. O. Twardziok,
Vera Vendramin, Simone Scalabrin, Davide Scaglione,
Giuseppe Sciara, Simona Corneti, Matteo Bozzoli, Andrea Massi,
Klaus F.X. Mayer, Curtis J. Pozniak, Michele Morgante, Roberto Tuberosa

Plant & Animal Genome XXVIII

January 11-15, 2020
San Diego, CA, USA

Objectives

- To investigate and provide a large database of SNPs among cultivated elite durum wheat varieties

- To assess the transcriptome diversity in elite durum wheat in relation to a high quality reference assembly of durum wheat cv. Svevo

- To analyse the gene expression variation based on varieties and tissues/organs using RNA-seq

- To investigate the gene loss/deletion during the polyploidisation events using the transcriptome of the 13 varieties in relation to bread wheat reference genome (cv. Chinese Spring IWGSC RefSeq) and genome assemblies of the 10+ Wheat Genomes Project

13 elite worldwide durum wheat cultivars (from 1940 to 2004 including diverse germplasm)

Accession	Year	Germplasm	Genotype feature
Capeiti 8	1940	Italian	Founder
Creso	1974	Italian/CIMMYT	Founder and parent of mapping population
Valnova	1975	Italian/North Am.	Founder
Edmore	1978	North Am.	Founder
Yavaros 79	1979	CIMMYT-'70	Founder
Altar 84	1984	CIMMYT-'80	Founder
Neodur	1987	French/North Am.	Parent of mapping population
Kofa	1996	Desert Durum	Parent of mapping population
Svevo	1996	Italian/CIMMYT	Parent of mapping population
Meridiano	1998	Italian	Parent of mapping population
Saragolla	2002	Italian/CIMMYT	Elite genotype
Claudio	2004	Italian/CIMMYT	Parent of mapping population
Strongfield	2004	North Am.	Elite genotype

RNA-seq of the 13 varieties

An in-depth wheat transcriptome obtained based on ILLUMINA RNA-seq from the following tissues/organs:

- Young roots
- Young leaves
- Grains at the developing stage

SNP database

Genomic tools for durum wheat breeding: de novo assembly of Svevo transcriptome and SNP discovery in elite germplasm

Vera Vendramin[†], Danara Ormanbekova[†], Simone Scalabrin, Davide Scaglione, Marco Maccaferri, Pierluigi Martelli, Silvio Salvi, Irena Jurman, Rita Casadio, Federica Cattonaro, Roberto Tuberosa, Andrea Massi and Michele Morgante

[†]Contributed equally

BMC Genomics 2019 20:278

https://doi.org/10.1186/s12864-019-5645-x © The Author(s). 2019

- Reference *de novo* assembled transcriptome sequence of cv. Svevo
- A high-quality and extensive SNP dataset of *T. durum*-specific SNPs
- SNPs identified among the 13 cultivars representing a wide range of genetic diversity widely diffused in the Mediterranean region and worldwide

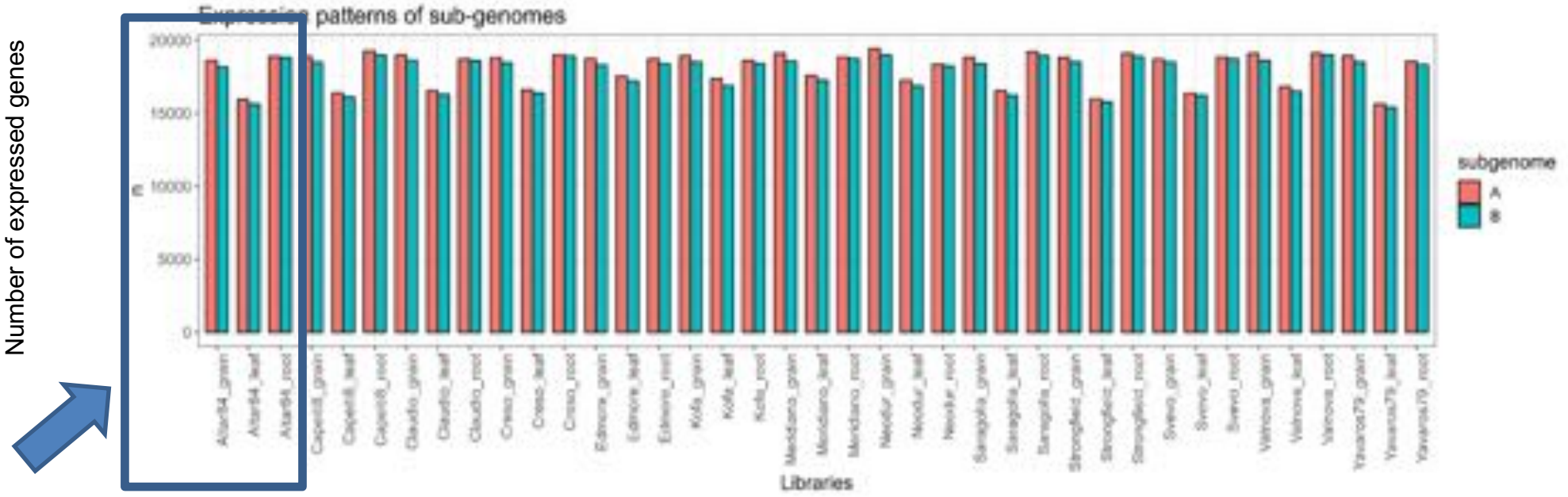
SNPs	#	%
Simple SNPs (single genome, diploid behavious)	33,747	5.4
Varietal Hemi-SNPs	497,783	79.4
SNPs present in more than two genotypes	**95,358**	15.2
Total SNPs computed	626,888	
Selected SNPs for the wheat 90K *iSelect* chip	**7,940**	

- 7,940 SNPs based on Illumina Assay Design Tool (ADT) score were selected to be included in the wheat 90K Illumina iSelect SNP array

Gene expression results

Out of 63,993 Svevo HC genes:

- **55,428 (86.6%) expressed in at least one tissue and/or variety**
- **48,007 (75.0%) expressed in grain tissues**
- **45,142 (70.5%) expressed in leaves**
- **47,702 (74.5%) expressed in roots**
- **approx. 2% higher number of genes expressed in sub-genome A *vs*. sub-genome B**



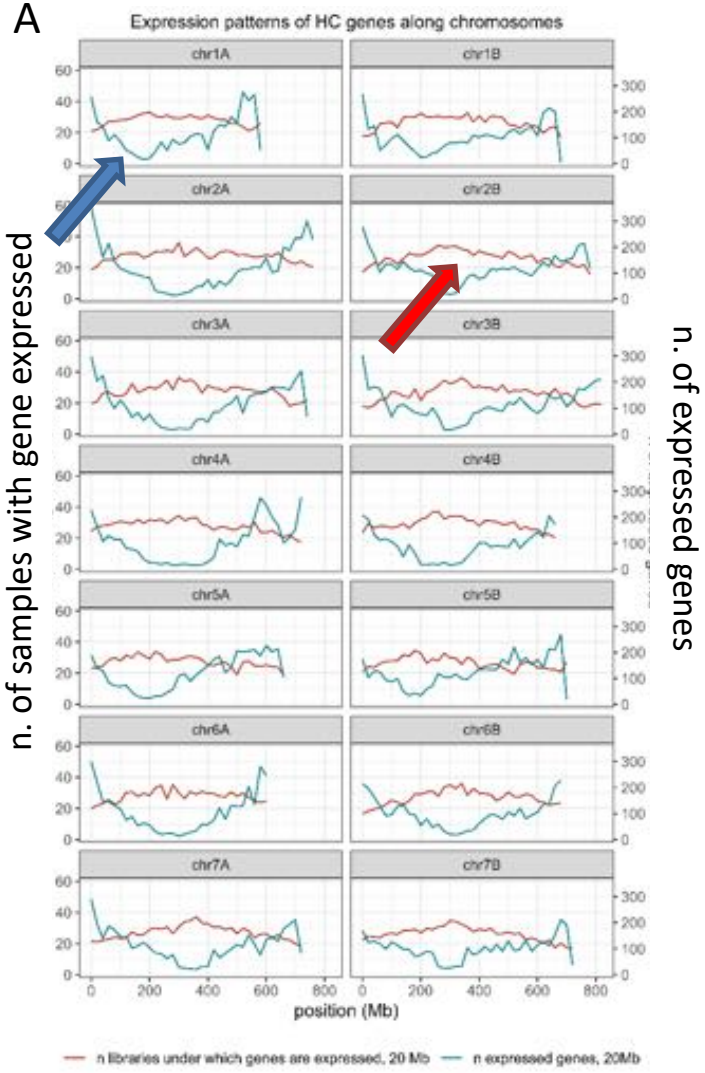
Gene expression pattern on chromosomes

n. of samples in which genes are expressed, 20 Mb window

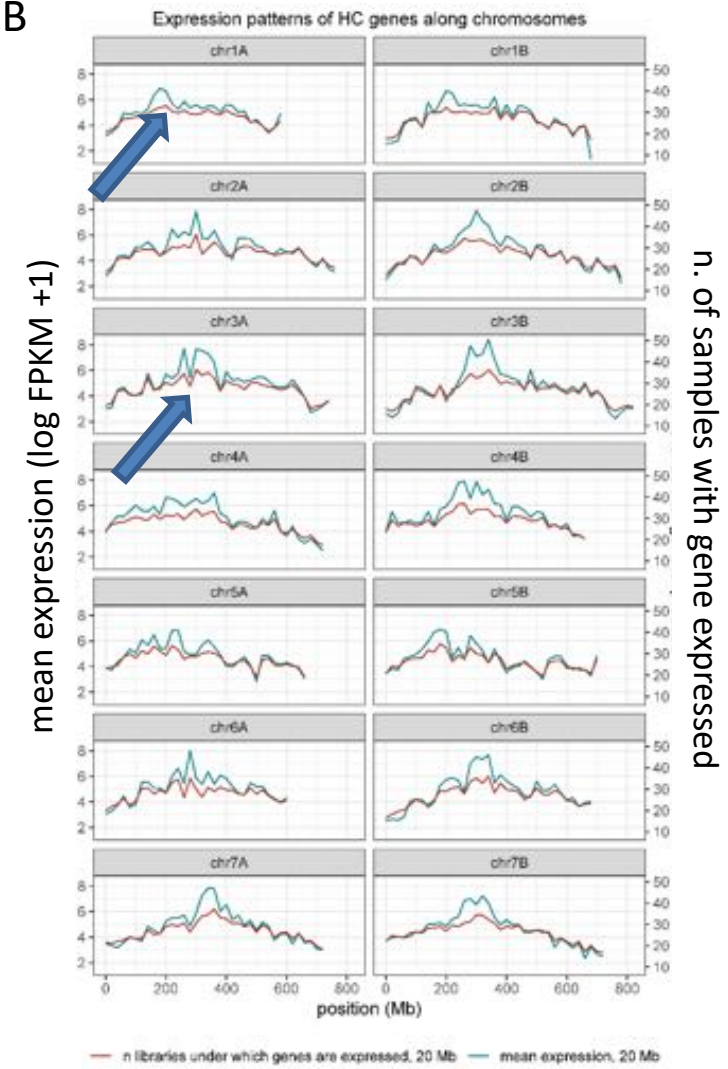
n. of expressed genes, 20 Mb

mean expression, 20 Mb

A



B

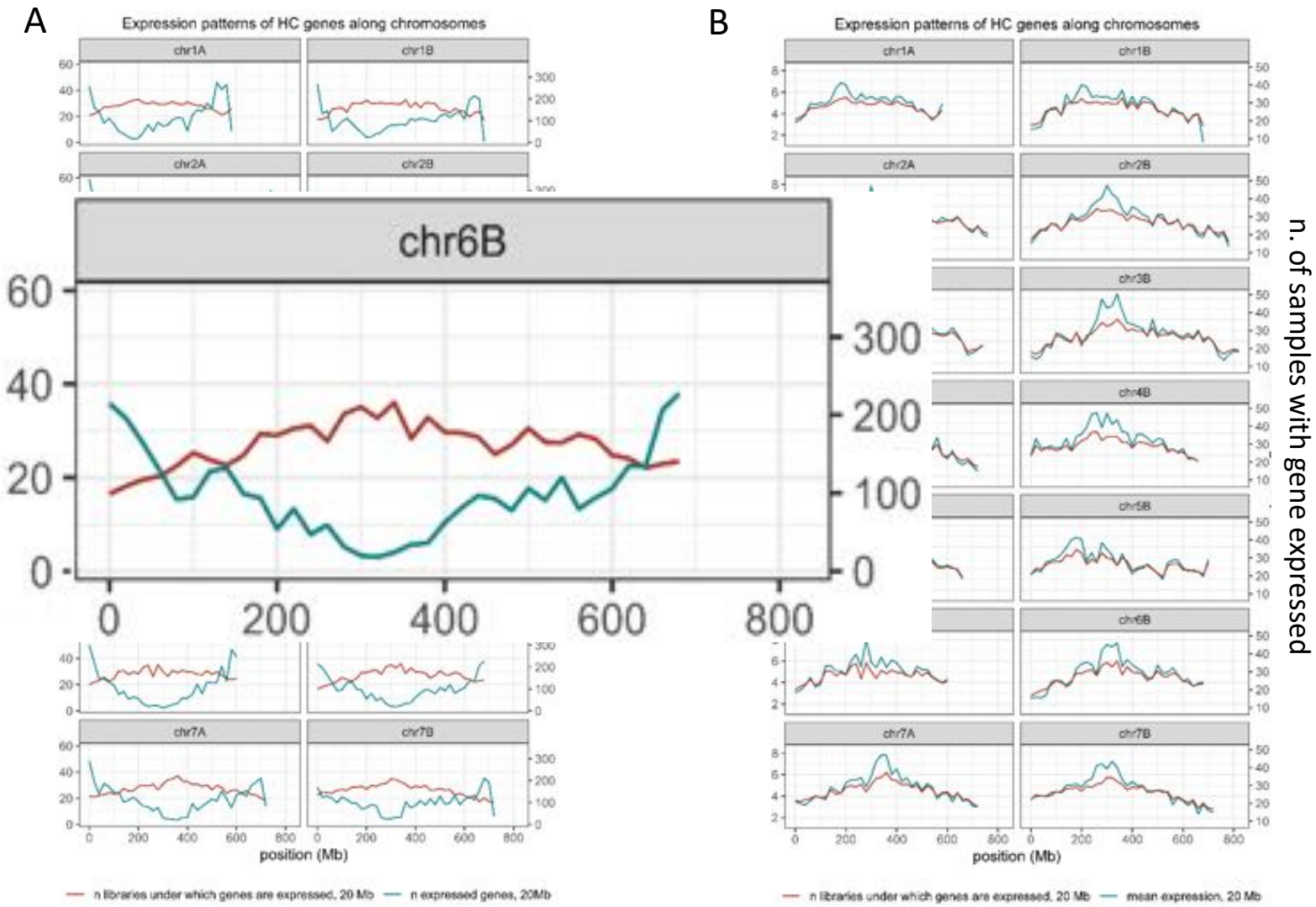


Gene expression pattern on chromosomes

n. of samples in which genes are expressed, 20 Mb window

n. of expressed genes, 20 Mb

mean expression, 20 Mb

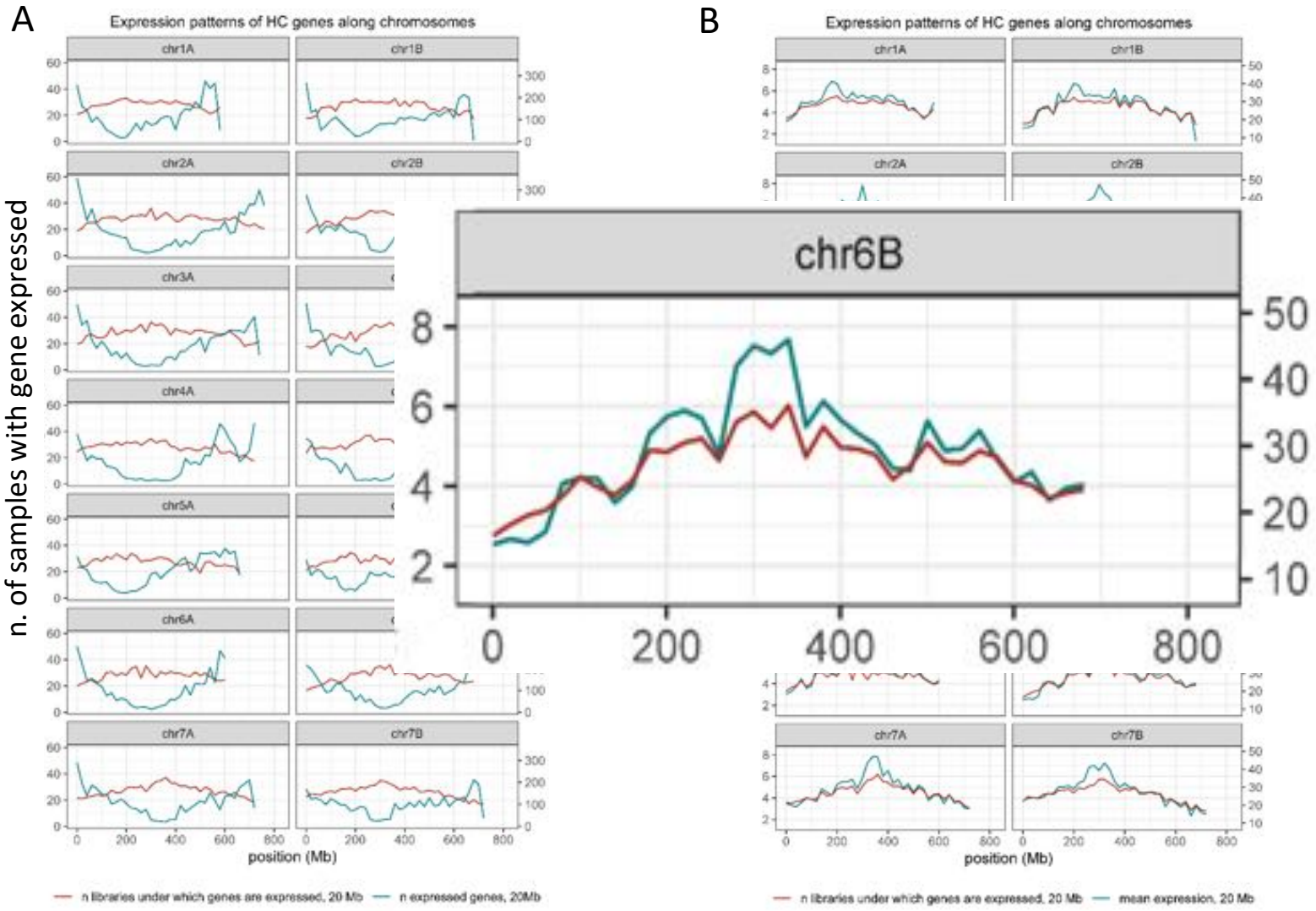


Gene expression pattern on chromosomes

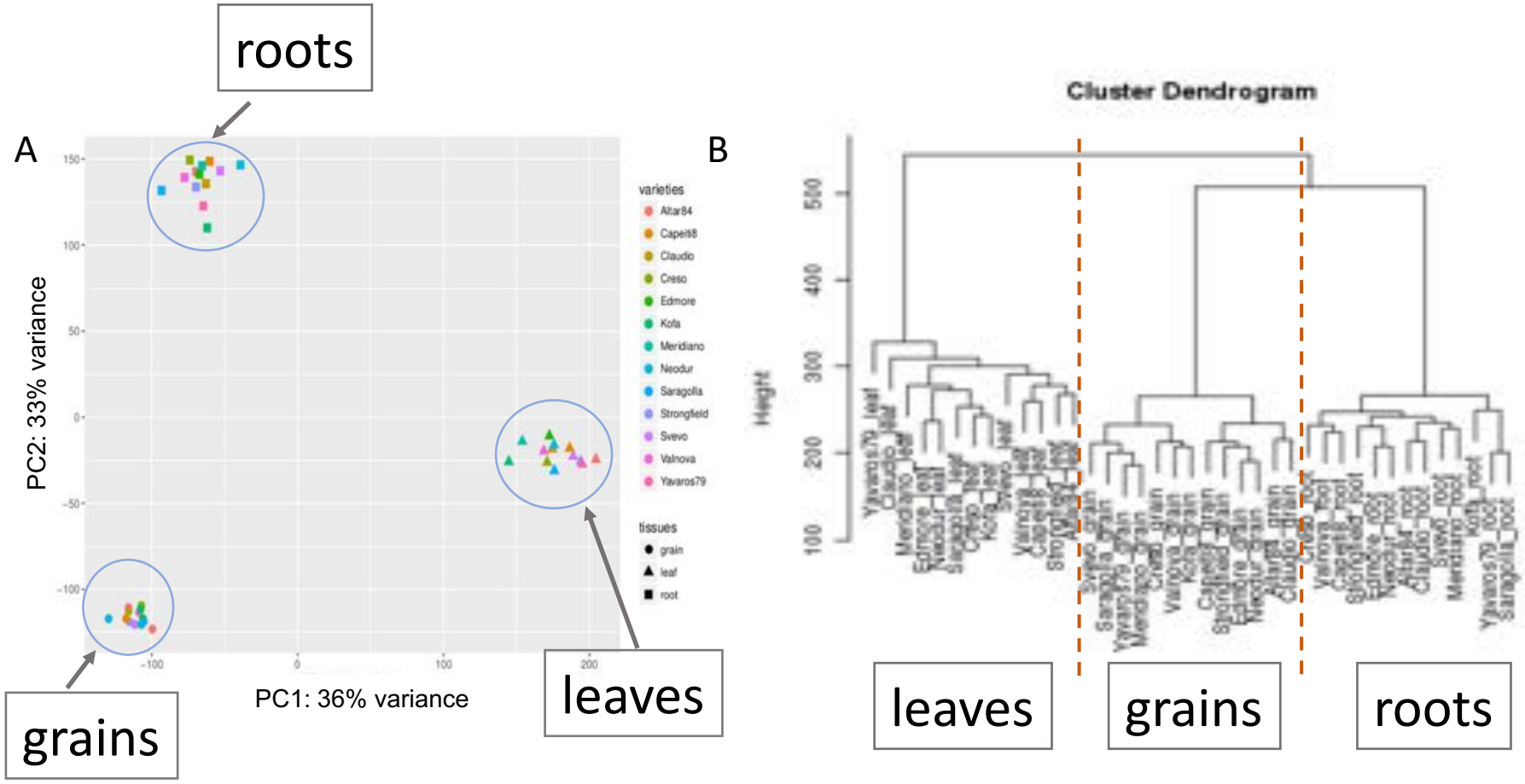
n. of samples in which genes are expressed, 20 Mb window

n. of expressed genes, 20 Mb

mean expression, 20 Mb



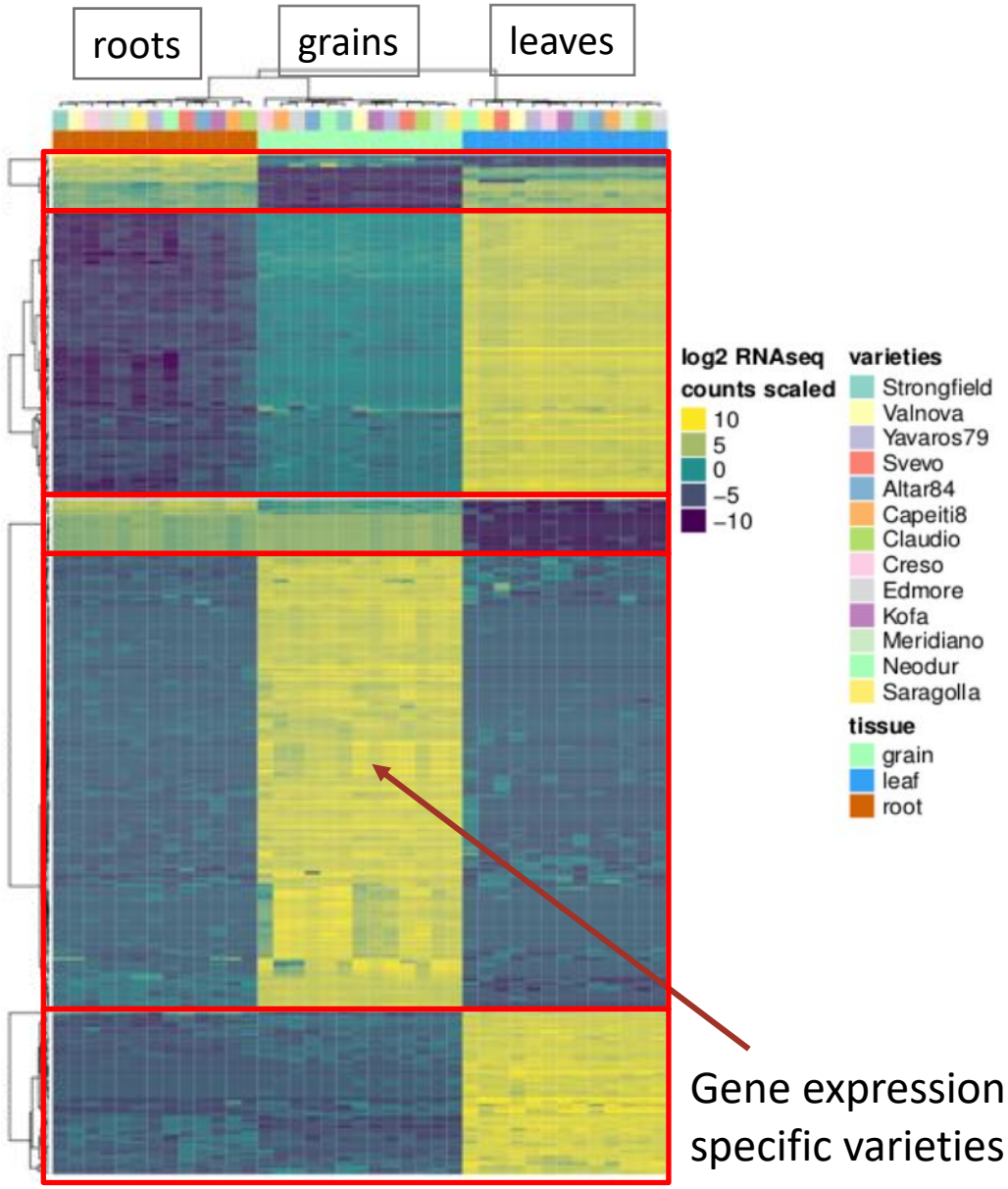
Principal Component Analysis and Hierarchical clustering: clustering lead by tissues



A) two-dimensional Principal Component Analysis (PCA). Clustering lead by tissues with 36 % variance (PC1) + 33 % variance (PC2).

B) hierarchical cluster analysis based on dissimilarities

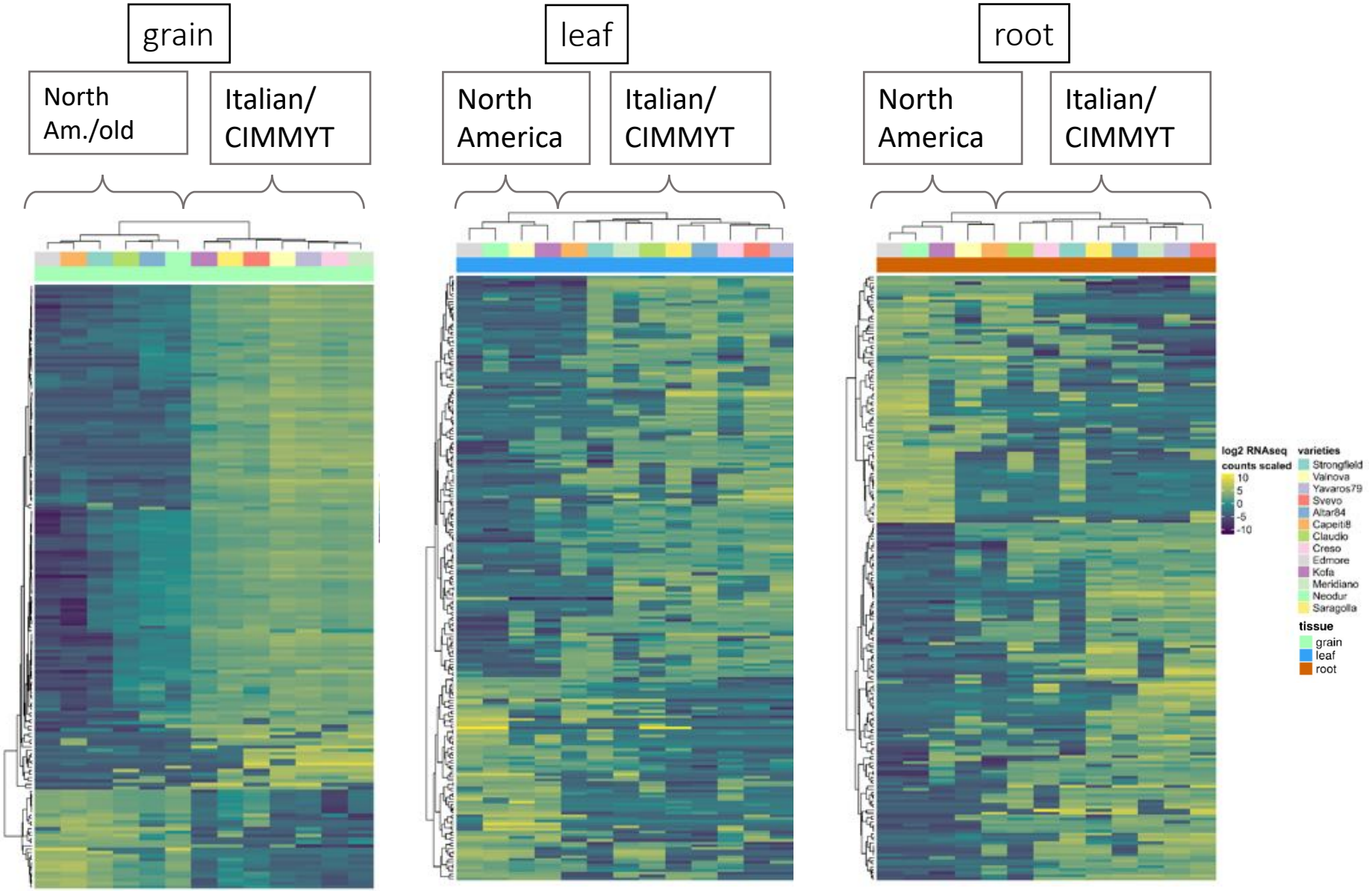
Hierarchical clustering based on sparse PCA: genes expression variation in tissues and varieties



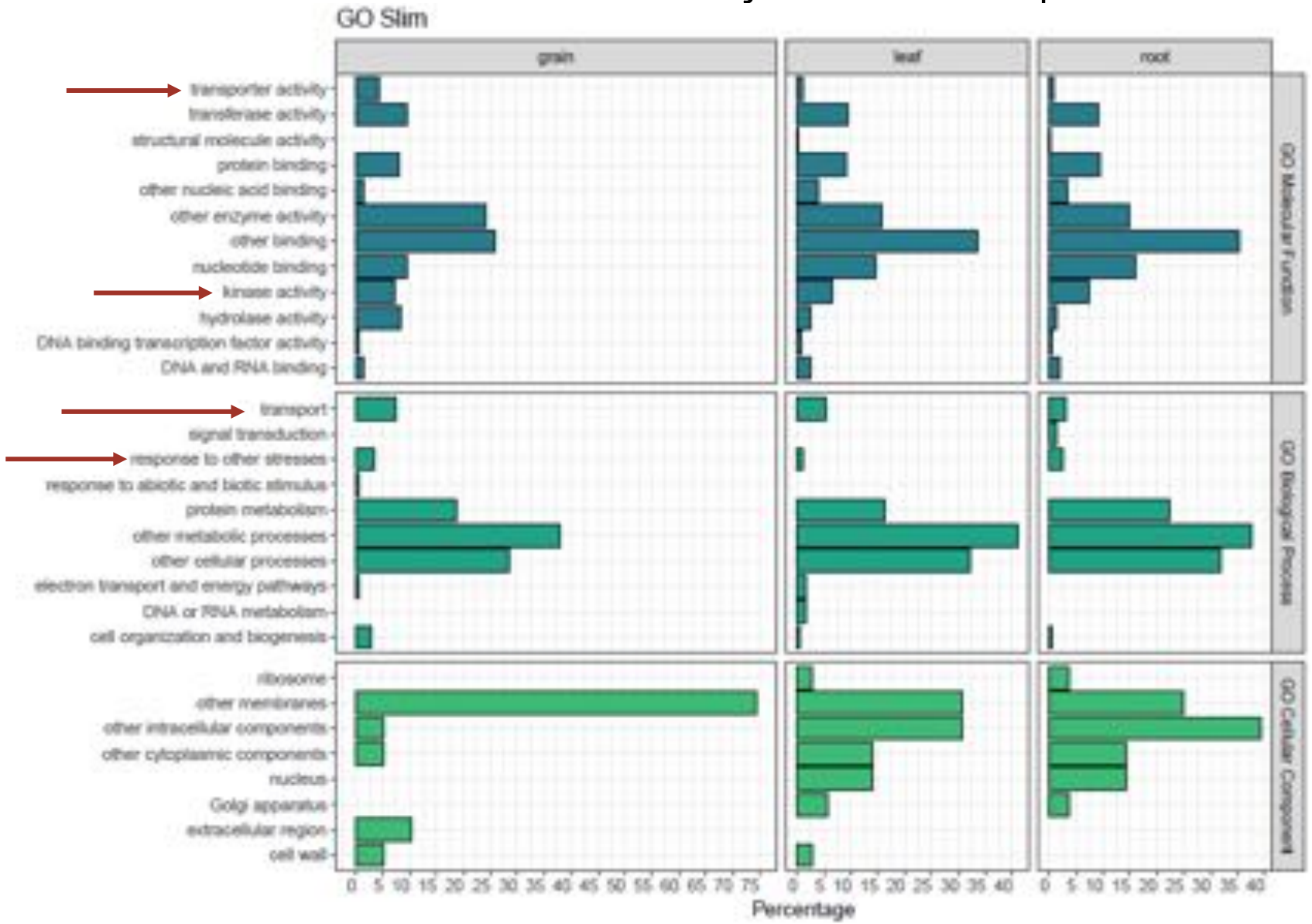
Gene expression clusters in specific varieties and tissues

- Sparse PCA based on the strongest PC1-PC2 scores
- 510 genes contributing to major variation patterns were retained (over 55,428 expressed genes cv. Svevo HC set)
- Clearly differentiated up- and down-regulated gene expression clusters based on tissues and varieties

Hierarchical clustering of cultivar expression profiles based on sparse PCA, **clustering led by genetic relationships**

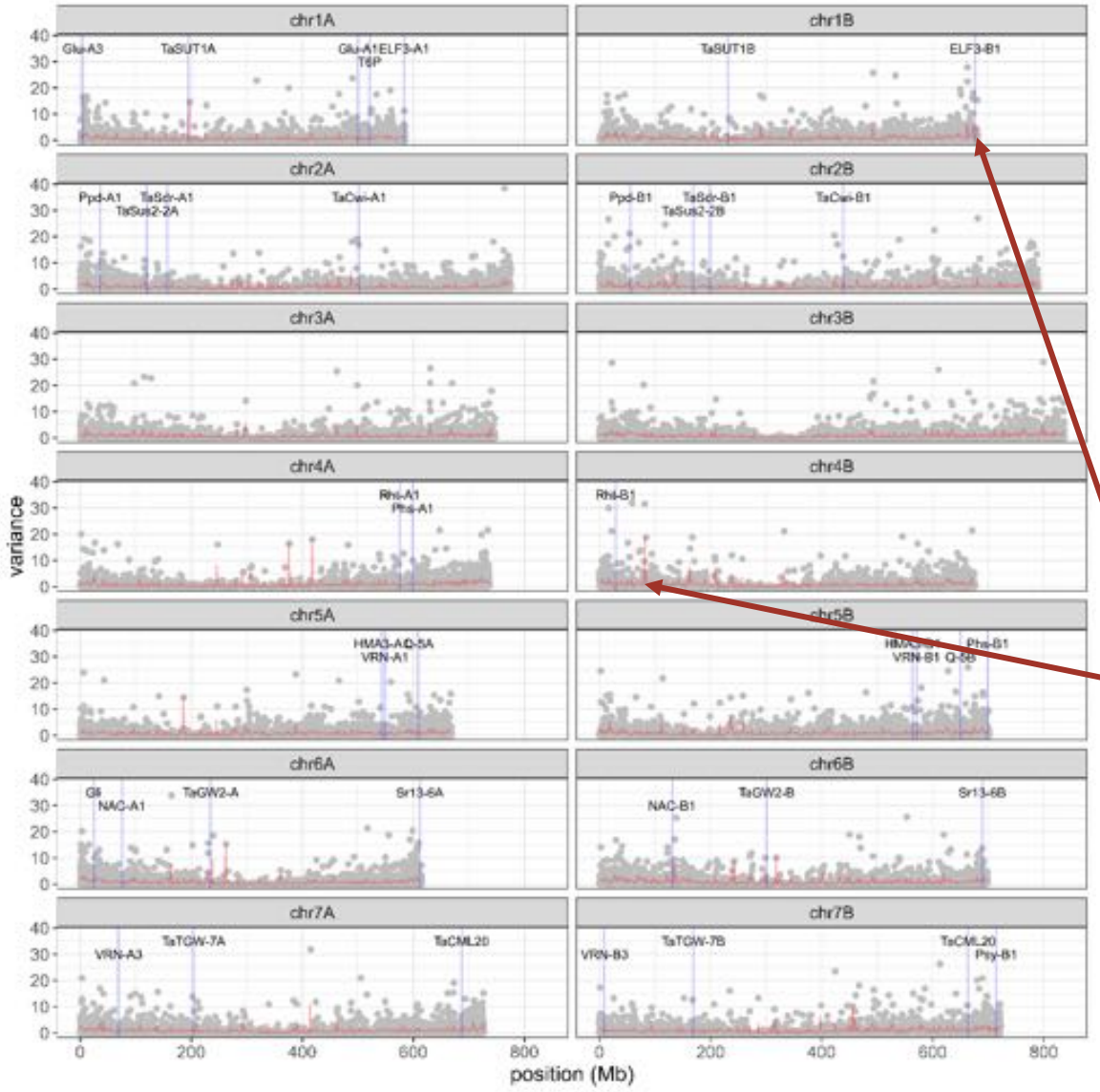


GO Slim Classification of genes contributing to major variation pattern



- 510 genes contributing to major variation patterns were classified using GO (Gene Onthology) Slim
- The majority of genes were involved in **kinase activity, transport, response to stresses**

Variance expression analysis of varieties for grain

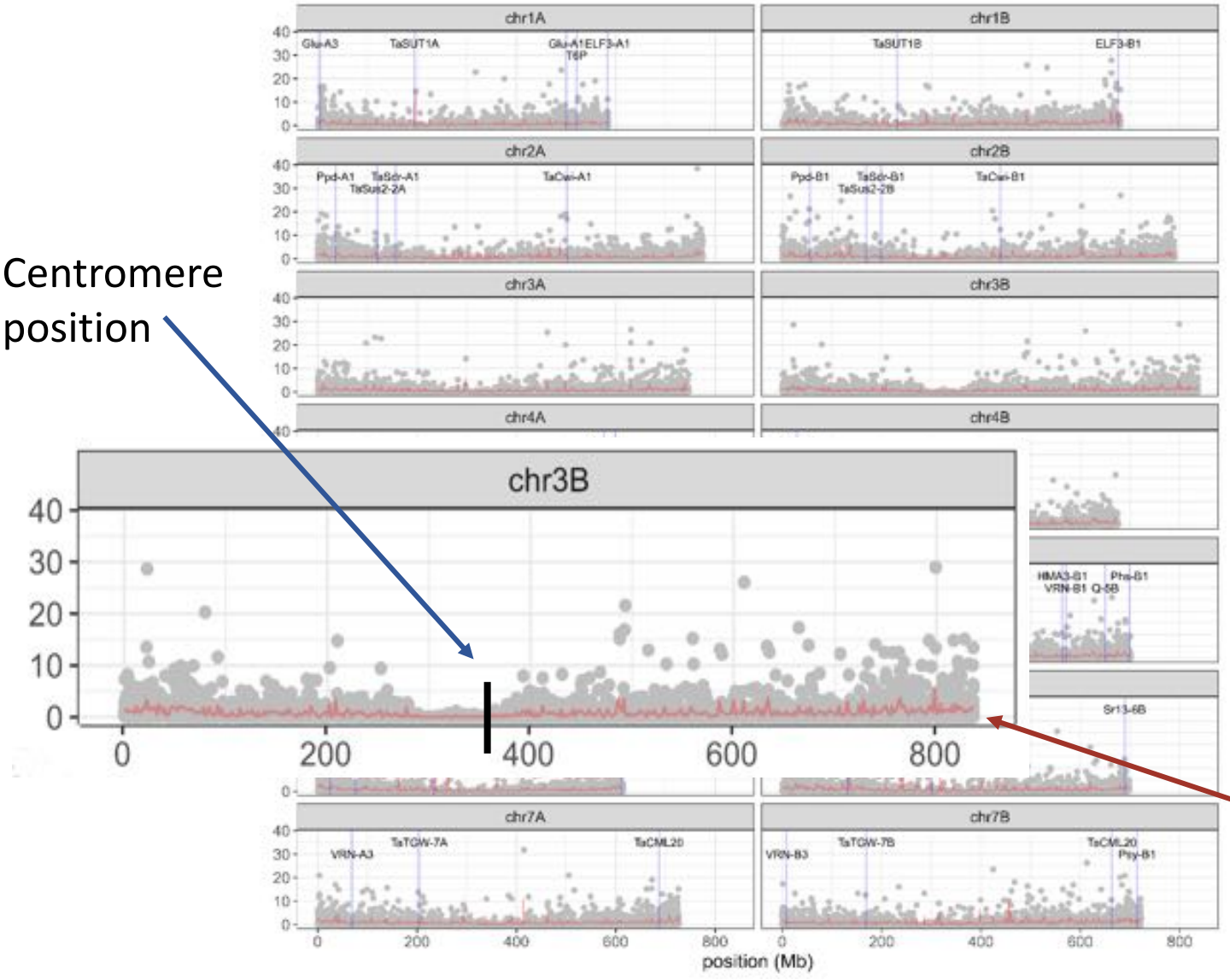


- Variance expression analysis projected on the Svevo genome assembly
- Allow to identify the chromosome regions that drive the major expression variation pattern

Red line indicates mean variance over 20 Mb window

Variance expression analysis of varieties for grain

Centromere position

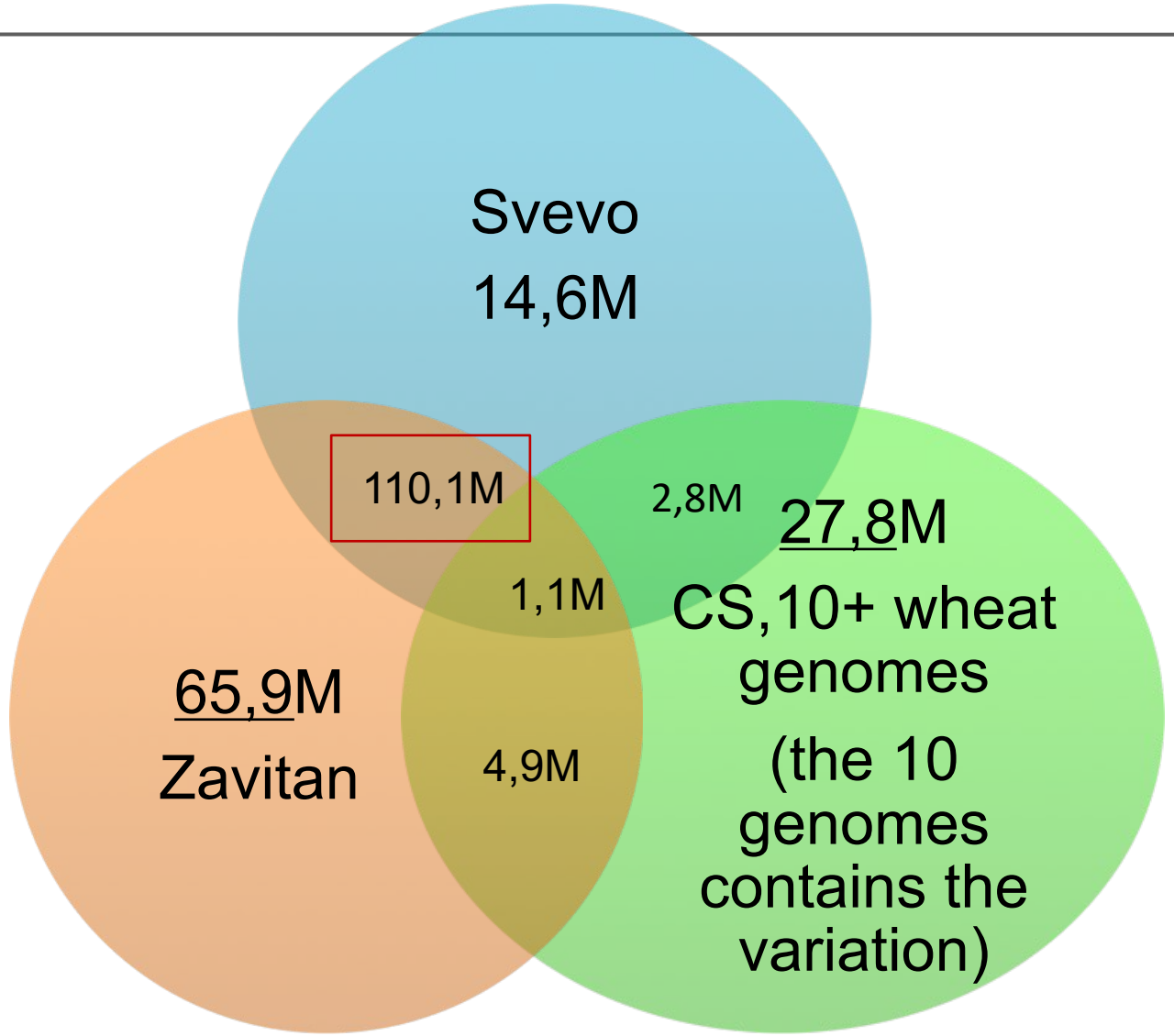


- Variance expression analysis projected on the Svevo genome assembly
- Allow to identify the chromosome regions that drive the major expression variation pattern

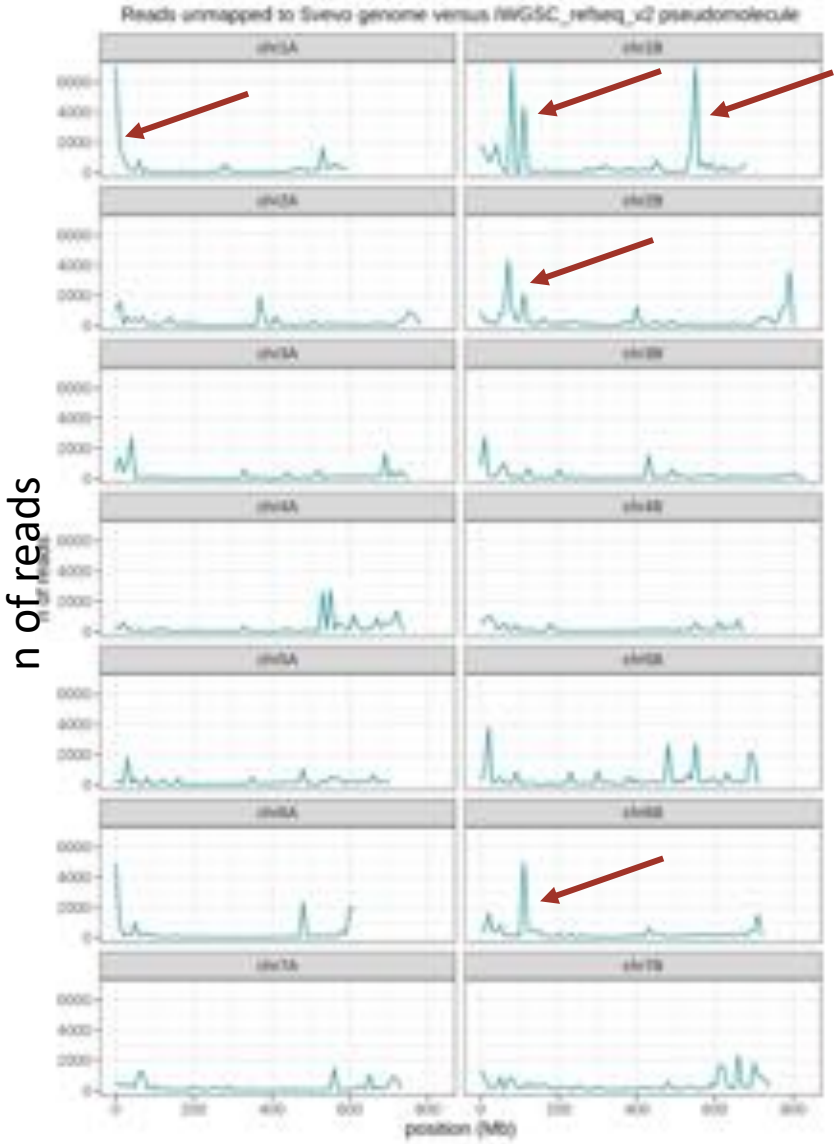
Red line indicates mean variance over 20 Mb window

Venn's Diagram of Unmapped reads

- RNA-seq reads of 13 durum wheat varieties were also mapped against:
  - IWGSC_refseq_v2
  - WEW_Zavitan_v1
- We identified the number of unmapped reads in common between the three genomes



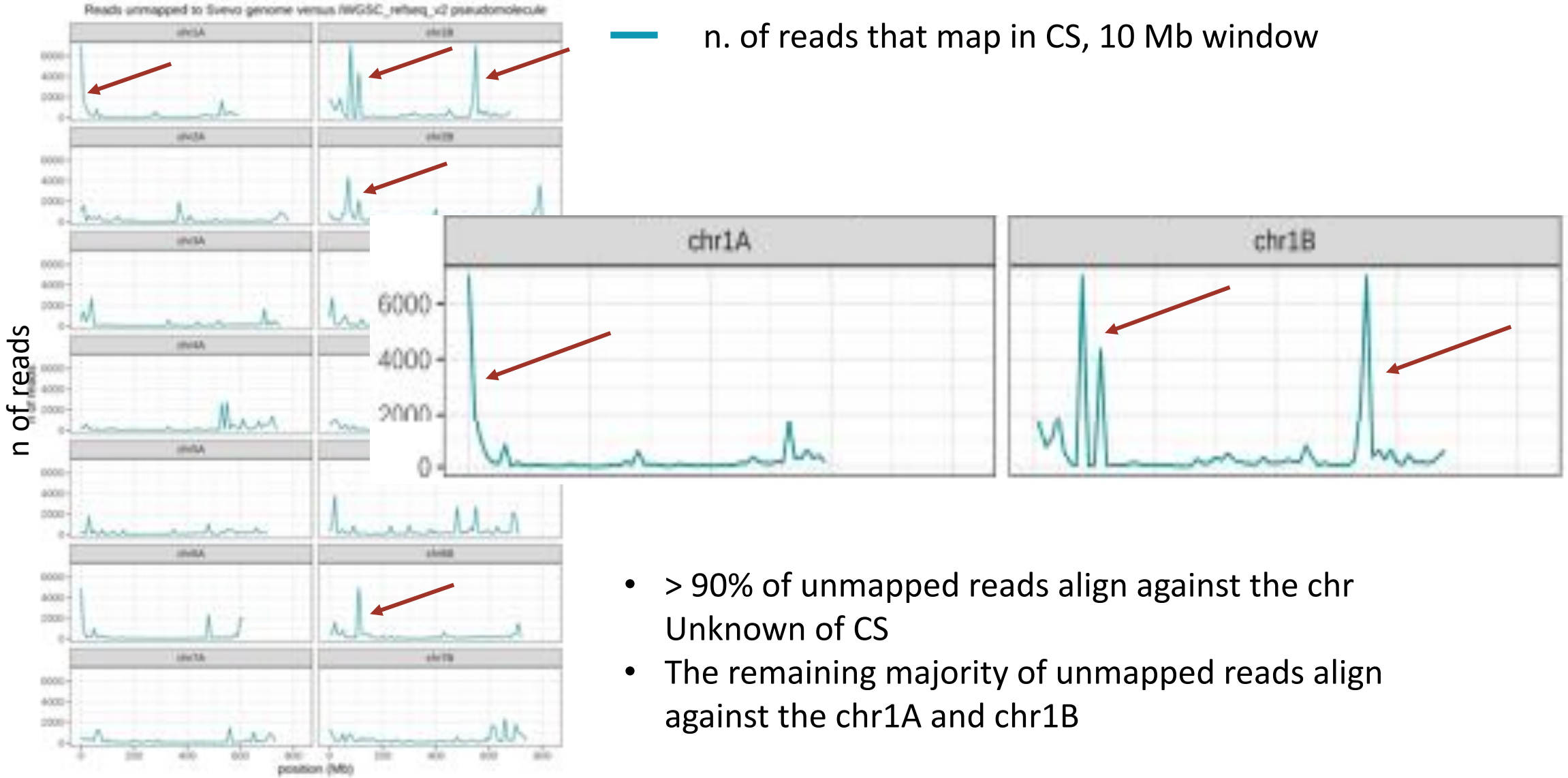
Reads unmapped to Svevo genome aligned against CS pseudomolecule



n. of reads that map in CS, 10 Mb window

- > 90% of unmapped reads align against the chr Unknown of CS
- The remaining majority of unmapped reads align against the chr1A and chr1B

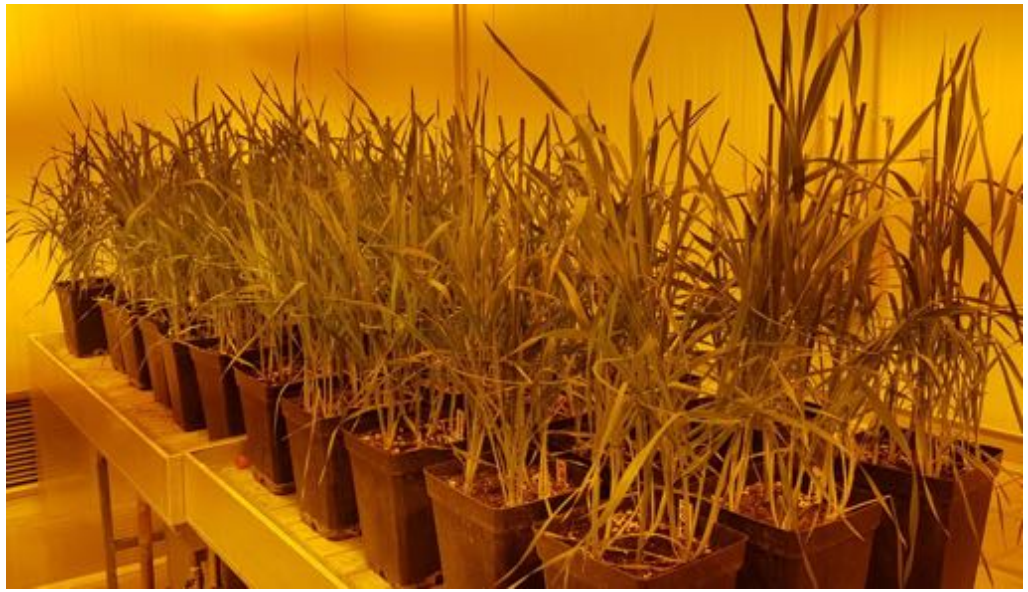
Reads unmapped to Svevo genome aligned against CS pseudomolecule



- > 90% of unmapped reads align against the chr Unknown of CS
- The remaining majority of unmapped reads align against the chr1A and chr1B

More tissues for RNA-seq

- To expand the experiment 13 varieties were re-grown in growth chamber
- **The following samples were collected:**
  - **ovaries and anthers at heading**
  - **three separated samples of grains at 3, 11 and 16 days after pollination corresponding to three critical stages of grain development up to the full synthesis of glutenin/gliadin proteins**



Further perspectives and concluding remarks

- **We have a pan-transciprome in durum, the cultivar specific reads that were not mapped on Svevo genome (4-30 %) are being *de novo* assembled**

- **The expression pattern database could be useful to identify genes regulated by eQTL and to elucidate the function of candidate gene**

- **Characterizing the gene expression presence-absence variation (ePAV) in tetraploid durum wheat show extensive presence of this variation**

- **Using the transcriptome of 13 varieties in relation to bread wheat reference genome (cv. Chinese Spring IWGSC RefSeq) can help to explore the gene loss/deletion during the polyploidisation events**

- **With the availability of the 10+ Wheat Genomes Project, which include cultivars that represent genetic diversity, we aim to investigate presence of strong allele fixation events (allopolyploidisation bottleneck).**

Acknowledgements

AGER project "From Seed to Pasta – Multidisciplinary approaches for a more sustainable and high quality durum wheat production"

University of Bologna:

Danara Ormanbekova
Roberto Tuberosa
Giuseppe Sciara
Simona Corneti
Matteo Bozzoli
Eder Licieri Groli

HelmholtzZentrum münchen
German Research Center for Environmental Health

Plant Genome and Systems Biology:

Sven O. Twardziok
Manuel Spannagl
Klaus F. X. Mayer

IGATech

IGA Technology Services:

Vera Vendramin
Simone Scalabrin
Michele Morgante

The International Durum Wheat Genome Sequencing Consortium

Andrea Massi