HIGH-QUALITY ASSEMBLY OF THE DURUM WHEAT GENOME CV. SVEVO

Luigi Cattivelli,
The International Durum Wheat Genome Sequencing Consortium

Durum wheat

- Durum wheat with a total production of about 35 million tons is the 10th most important crops worldwide.
- Durum wheat is a tetraploid wheat with 12 Gb genome (80% repetive elements), domesticated from wild emmer in the Fertile Crescent about 10,000 years ago.
- The domestication process and the subsequent breeding activity have selected the modern durum wheat fixing the typical domestication traits (e.g. brittle rachis, naked seeds and threshability), agronomic traits (e.g. short straw and disease resistances) as well as specific pasta quality traits (grain hardness, pigment content, protein content).

An international initiative to sequence the genome of the durum wheat

crea
Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

Luigi Cattivelli
P. Faccioli, C Crosatti
Nicola Pecchioni

Consiglio Nazionale delle Ricerche

Aldo Ceriotti
Luciano Milanesi
Gabriella Sonnante

University of Bologna
M. Maccaferri, S.Salvi,
Roberto Tuberosa

Curtis Pozniak

IPK GATERSLEBEN

Nils Stein
Martin Mascher

MONTANA STATE UNIVERSITY

Hikmet Budak

אוניברסיטת תל-אביב

University of Tel Aviv
Assaf Distelfeld

HelmholtzZentrum münchen
German Research Center for Environmental Health

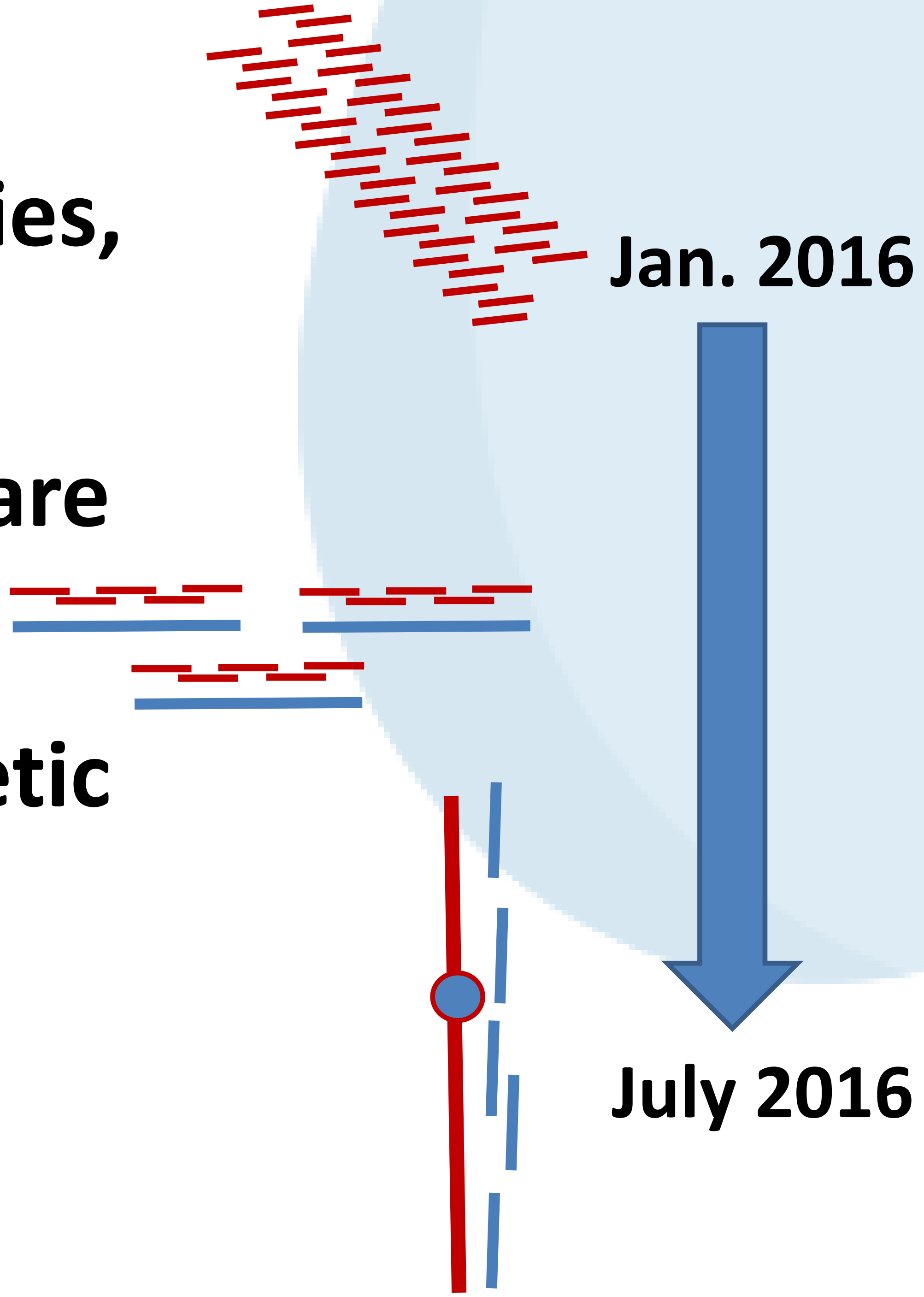
Klaus Mayer
Sven Twardziok
Heidrun Gundlach

Svevo, the first high-quality durum wheat

- **Svevo (1996): CIMMYT line × Zenit**
- **High protein content and high yellow pigment**

Sequencing the durum wheat variety SVEVO

- **Whole genome shotgun with Illumina technology: pair end and mate pair libraries, about 270x coverage.**
- **Assembly with the DeNovoMAGIC software by NRGene.**
- **Scaffolds anchored to a high-density genetic map (Svevo x Zavitan), and merged into super-scaffolds using Chromosome conformation capture (Hi-C).**
- **Genome annotations and analysis.**



High quality assembly of the Svevo genome: main statistics

- **The assembly has produced 129,464 scaffolds covering a total of 10.45 Gb.**
- **50% of the genome has been assembled in 493 scaffolds, with the smallest scaffold of 5.97 Mb.**
- **90% of the genome has been assembled in 2,019 scaffolds, with the smallest scaffold of 1.09 Mb.**

Total scaffolds	129,464
Assembly size	10.45 Gb
Gaps size	149 Mb
Gaps	1.42%
L50 (length)	5.97 Mb
N50 (sequences)	493
L90 (length)	1.09 Mb
N90 (sequences)	2,019

Anchoring the scaffolds to the chromosomes

- **95.3% of the scaffolds has been placed along the chromosomes and 90.0% has been oriented.**

Total scaffolds	129,464
Assembly size	10.45 Gb
Gaps size	149 Mb
Gaps	1.42%
L50 (length)	5.97 Mb
N50 (sequences)	493
L90 (length)	1.09 Mb
N90 (sequences)	2,019

Durum wheat genome annotation

- **The durum wheat assembly has been annotated for genes, small RNAs and transposable elements.**
- **Genes and miRNAs annotation were based on:**
  - **extensive transcriptomic analysis from Svevo based on more than 60 treatment x tissue combinations.**
  - ***Triticum diccocoides* samples (provided by A. Distelfeld)**
  - ***Triticum aestivum* CS samples (published data sets)**
  - **Annotated protein sequences from grass species**
  - ***Triticeae* protein sequences from Uniprot**
  - ***Triticeae* full length cDNA sequences (TriFLDB)**
  - **CS PacBio Iso-Seq data from leaf, root, seed, seedling, spike, stem (published by Earlham Institute)**

Transcript confidence classification

High confidence transcripts 66,559

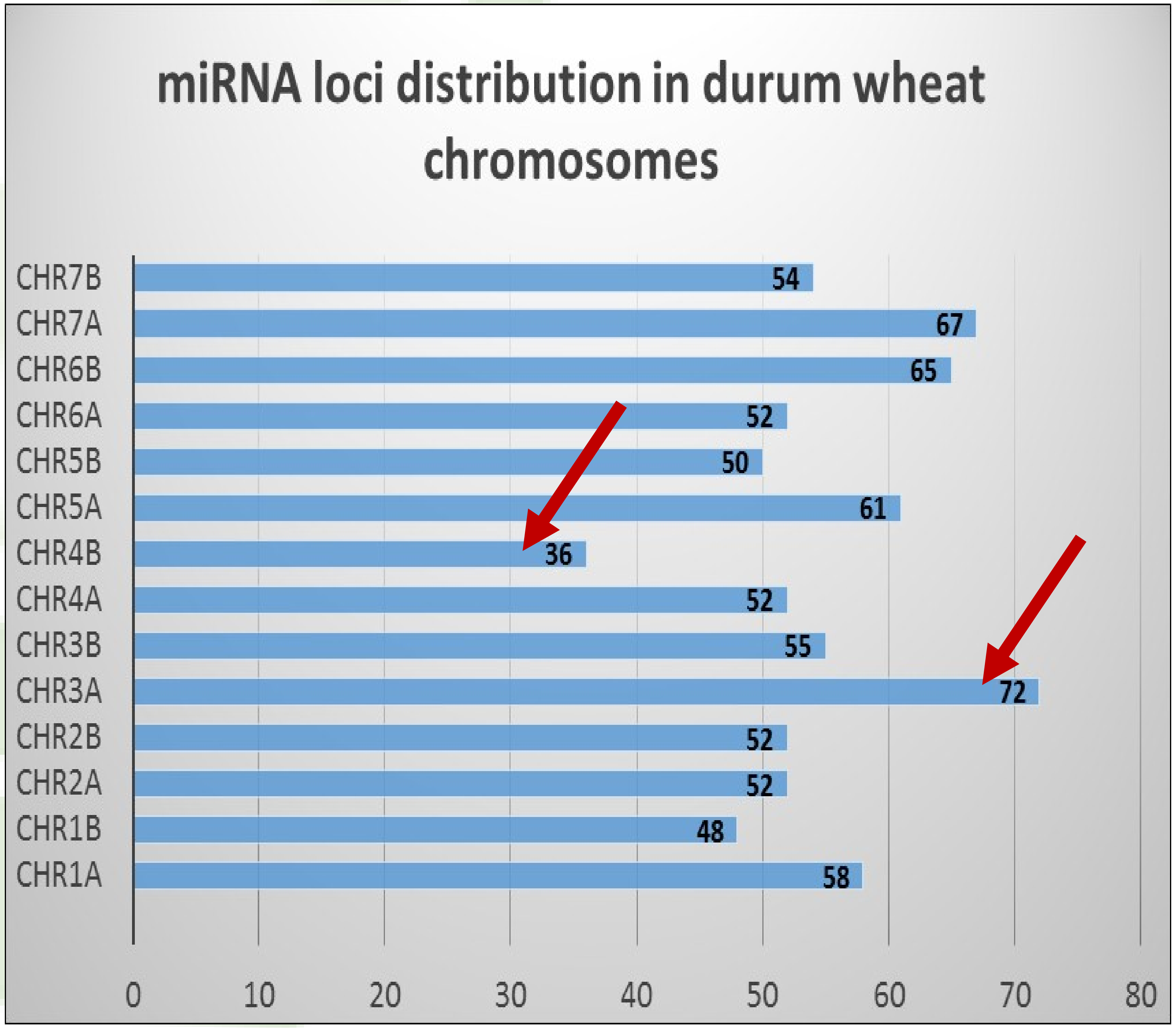
Coding sequence with annotated start and stop codon and significant match to reference protein sequence (query coverage > 90 % and subject coverage > 90 % and e-value < 10^{-10})

- ***HC1*: match to validated protein sequence (*Magnoliophyta*)**
- ***HC2*: match to predicted protein sequence (*Poaceae*)**

Low confidence transcripts 303,404

- ***LC1*: incomplete coding sequence but significant match to reference protein sequence**
- ***LC2*: no significant match to reference protein sequence but complete coding sequence**
- ***REP*: match to transposon elements database**

Durum wheat genome annotation: miRNAs



774 putative miRNA coding loci have been identified based on expression data

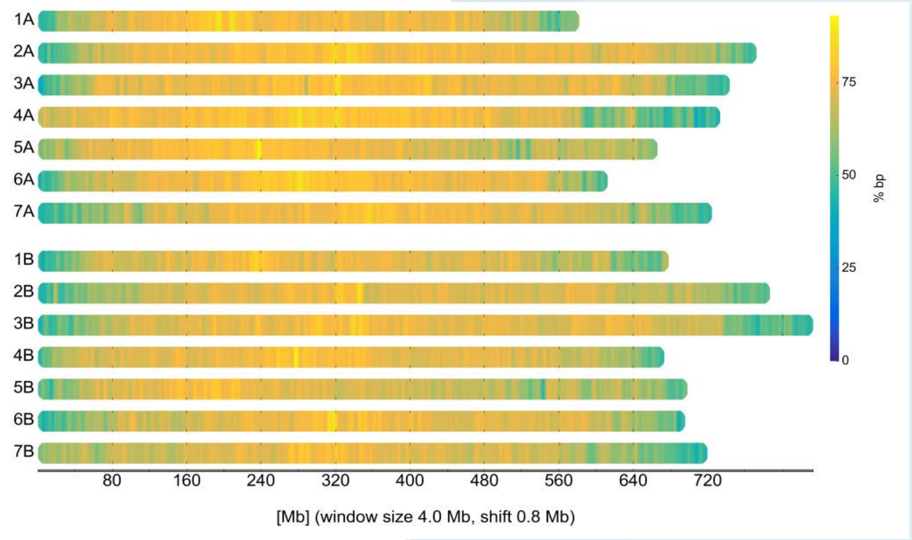
438 are homologous to annotated plant miRNAs and correspond to 79 MIR families

[Blast precursors vs miRBase v.21 (only plants) threshold: best hit, e-value 1.0e-5]

Durum wheat genome annotation: transposable elements

	% of genome	Mb
Mobile Element	**81.73**	**8,419**
Class I: Retroelement	**70.32**	**7,244**
Class II: DNA Transposon	**11.40**	**1,174**

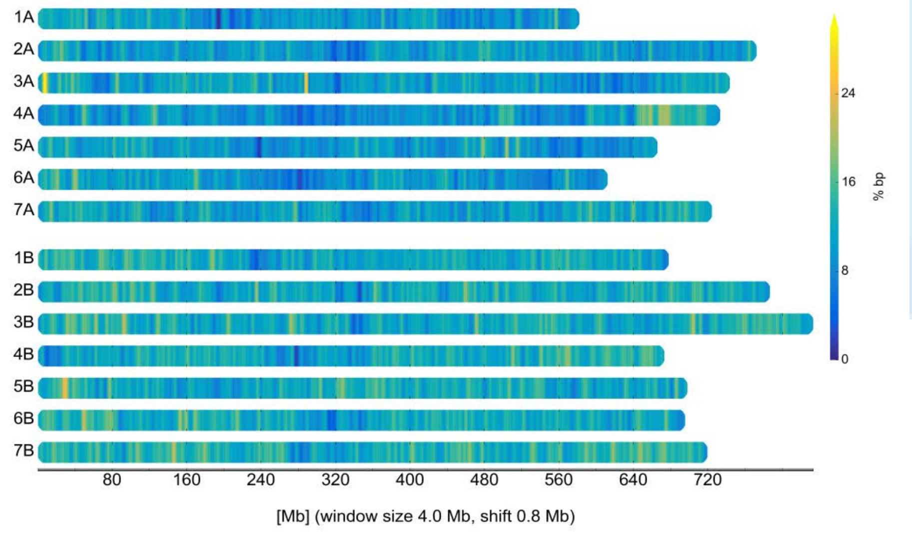
LTR-retrotransposons



A
71.5%

B
69.0%

DNA-transposons



A
10.3%

B
12.4%

B genome has some more DNA-TEs and slightly less retrotransposons

Conclusions

- **We provide a high quality assembly of the genome of the durum wheat cultivar Svevo, a contribution to the wheat pangenome analysis.**
- **Since the wild emmer wheat (acc. Zavitan) has been recently sequenced by A. Distelfeld, the comparison between wild emmer and durum wheat genomes (same high quality level) will shed light on wheat domestication and on selective sweeps (pasta quality trait).**
- **More information of the sequence of durum wheat genome in the presentation of Marco Maccaferri** (*Leveraging the assembly of the durum wheat genome*) **at the NRGene workshop on Tuesday 17h, 2.10pm.**

спасибо 谢谢

GRACIAS

THANK YOU

ありがとうございました MERCI

DANKE धन्यवाद

شُكراً OBRIGADO