

**HelmholtzZentrum münchen**

German Research Center for Environmental Health

**PGSB** Plant Genome and  
Systems Biology



# How complex can it be? The bread wheat genome and its implications for wheat intolerance

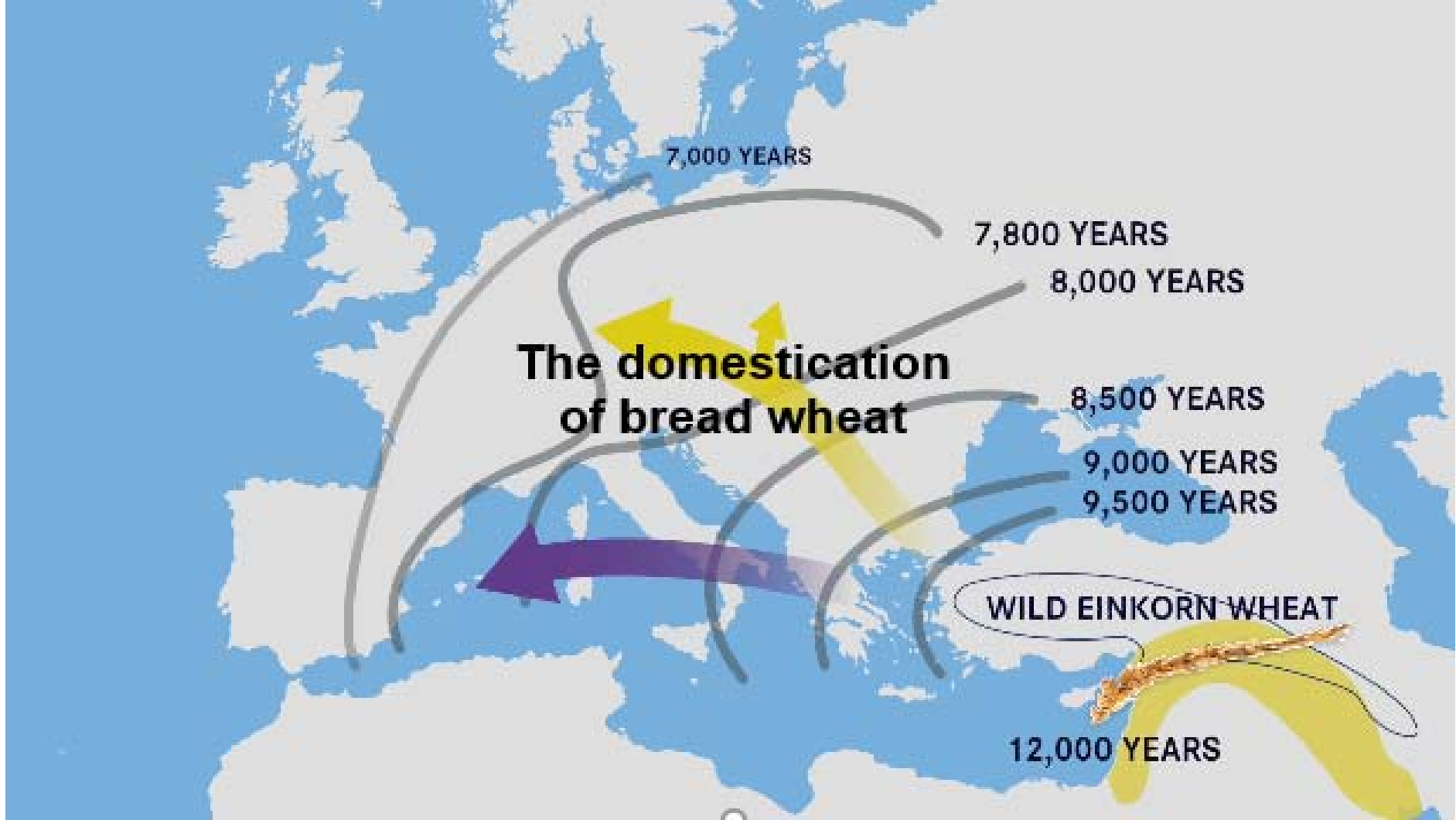
Manuel Spannagl

PGSB, March, 29<sup>th</sup> 2019

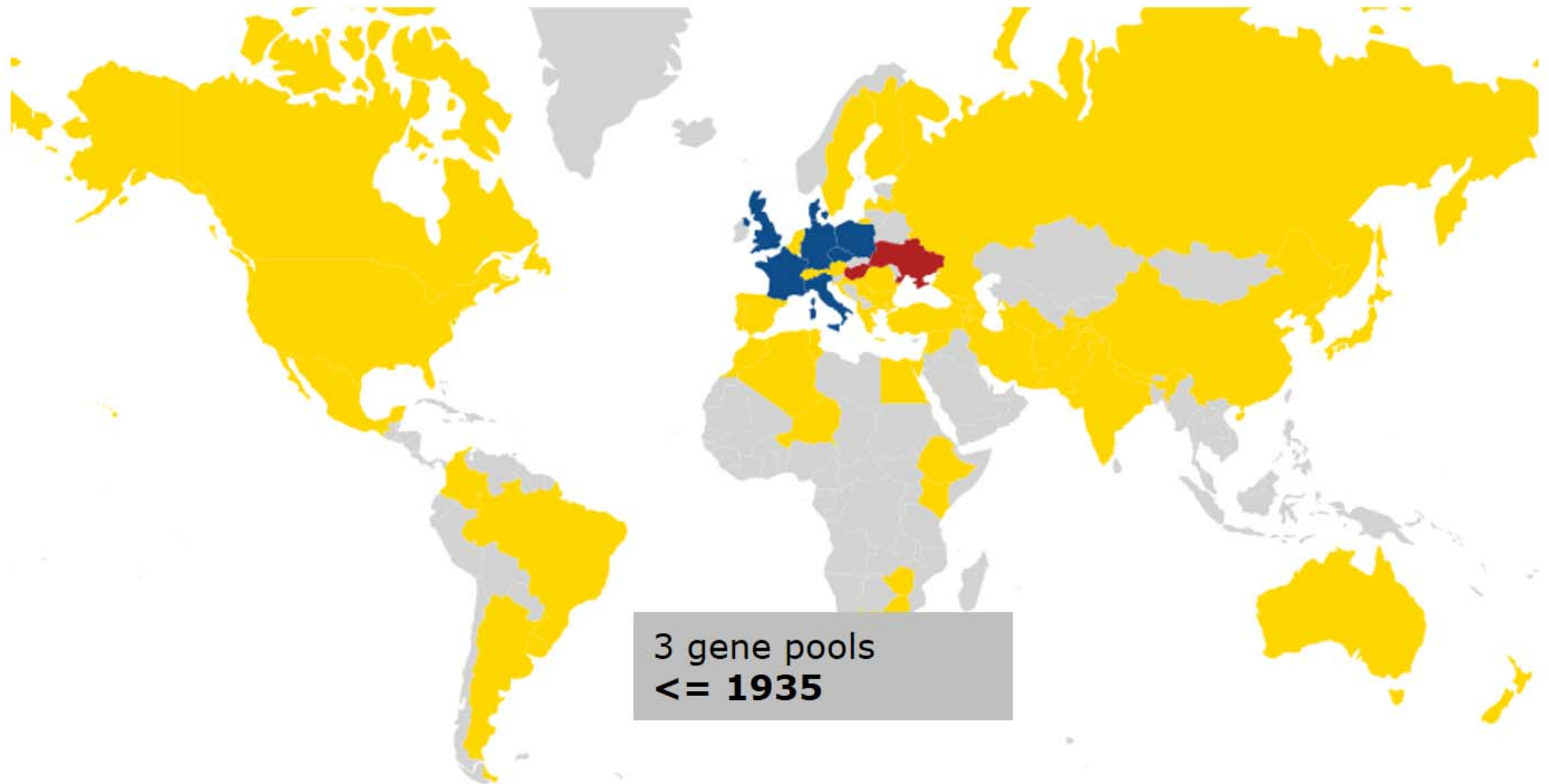
# Outline

- Introduction – what makes the wheat genome so special aka complex?
- Previous wheat WGS/survey sequences
- The bread wheat reference genome sequence
- The Wheat prolamin map
- What's next?





# Humans & even politics shape the gene pools of wheat



Nature Genetics, 2019



**Wheat (yield) is threatened  
by climate change**



# Wheat is not good for everyone



Celiac disease

Wheat insensitivities

Wheat allergy

Baker's asthma

# >60% of world's food:

Maize



Rice



Soy

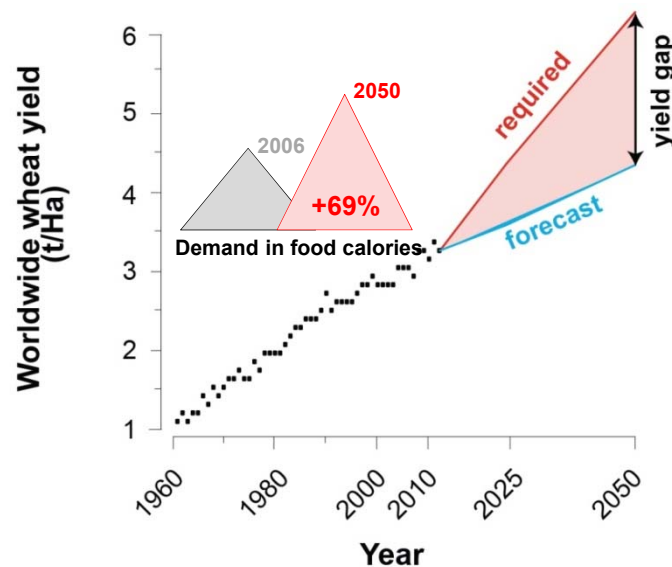


Wheat



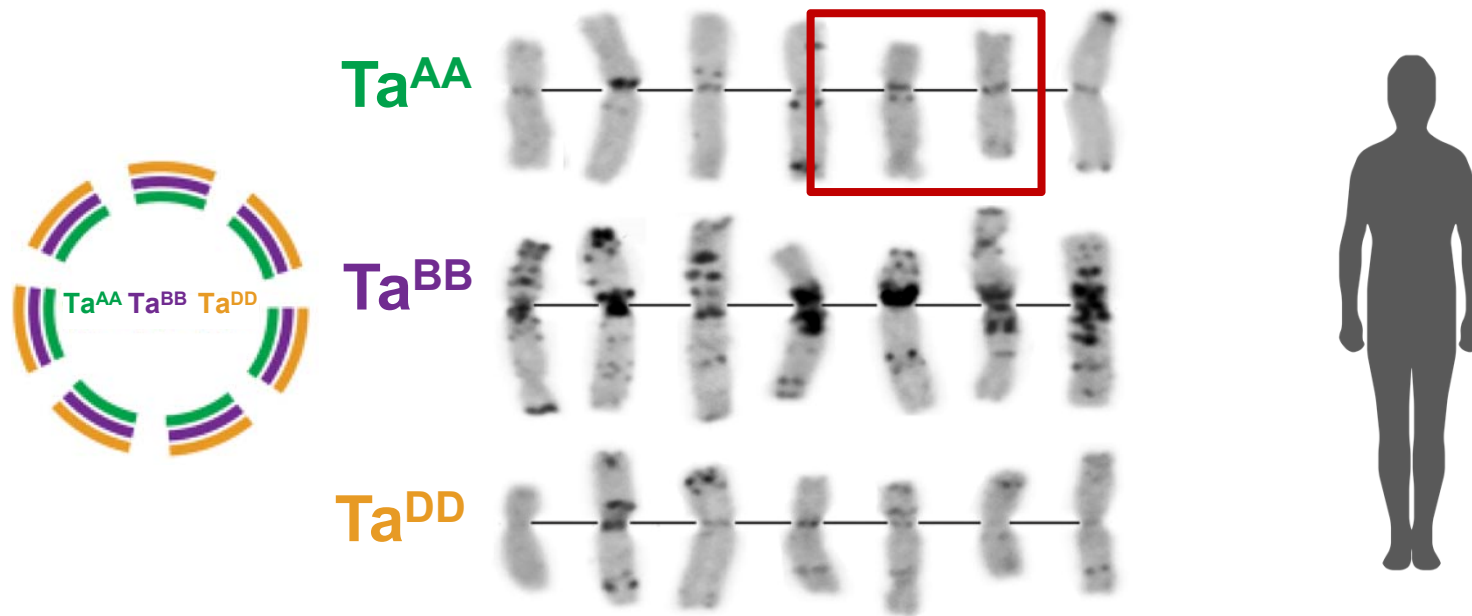
# Filling the yield gap – a challenge for agricultural research

- Staple food source for 30% of the world's population
- ~20% of human's daily consumed calories
- 620 million tonnes harvested (2012)
- US\$200 billion trade volume





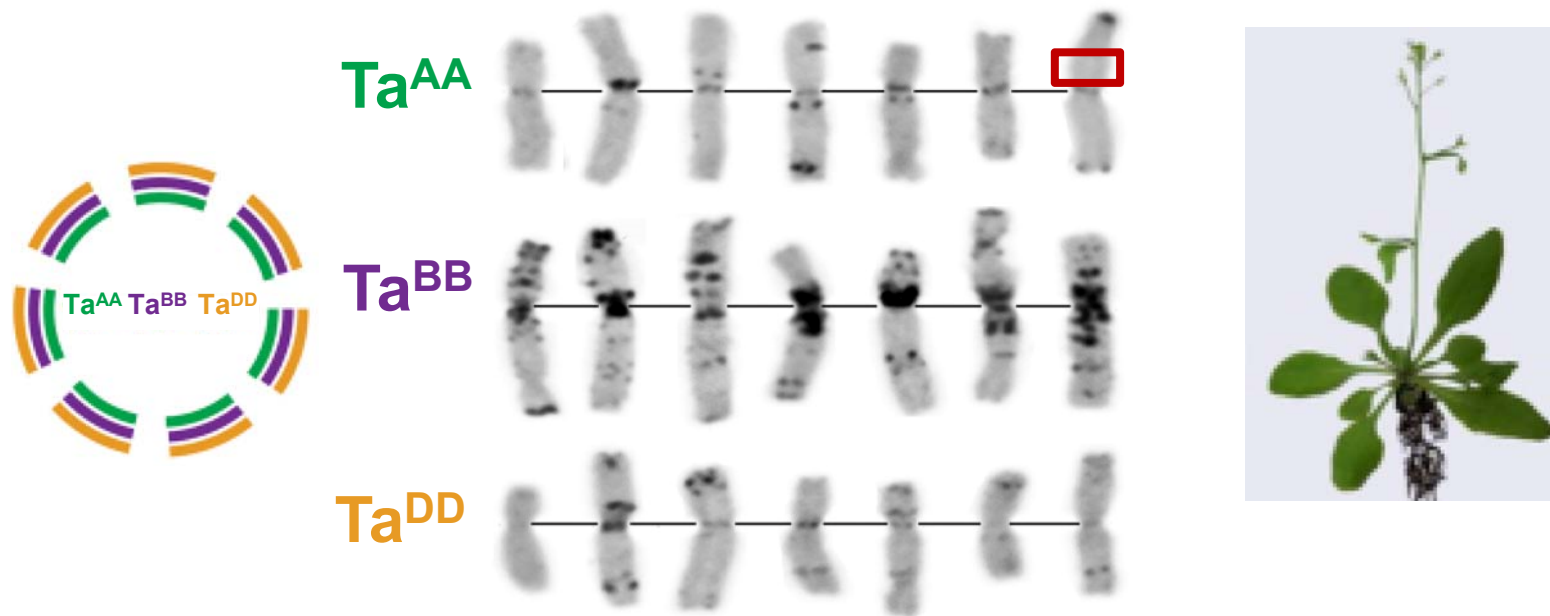
# The genome of bread wheat (*T. aestivum*) is ...



...large and complex!

- allohexaploid: three “diploid-like” ( $2n = 6x = 42$ )
- ~17Gb genome size; > 80% repeat content
- Highly similar genomes (>97% identity)

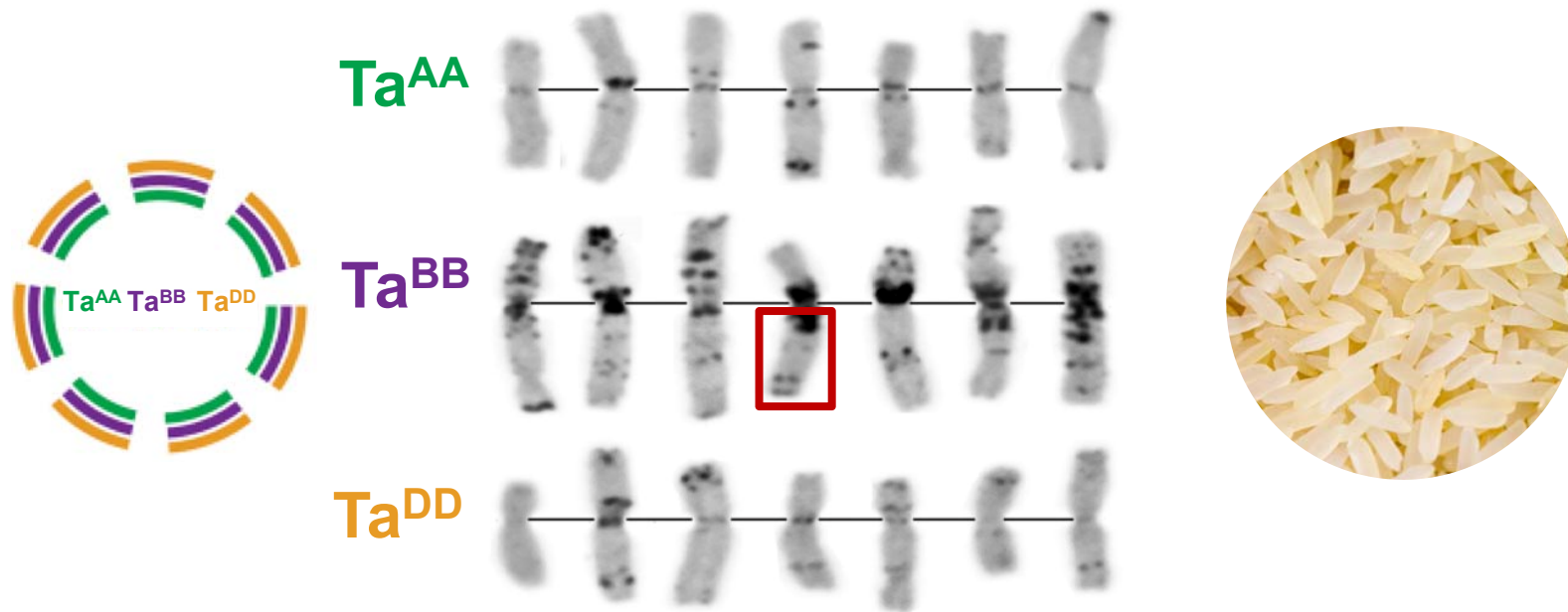
# The genome of bread wheat (*T. aestivum*) is ...



...large and complex!

- allohexaploid: three “diploid-like” ( $2n = 6x = 42$ )
- ~**17Gb** genome size; > **80% repeat** content
- Highly **similar genomes** (>97% identity)

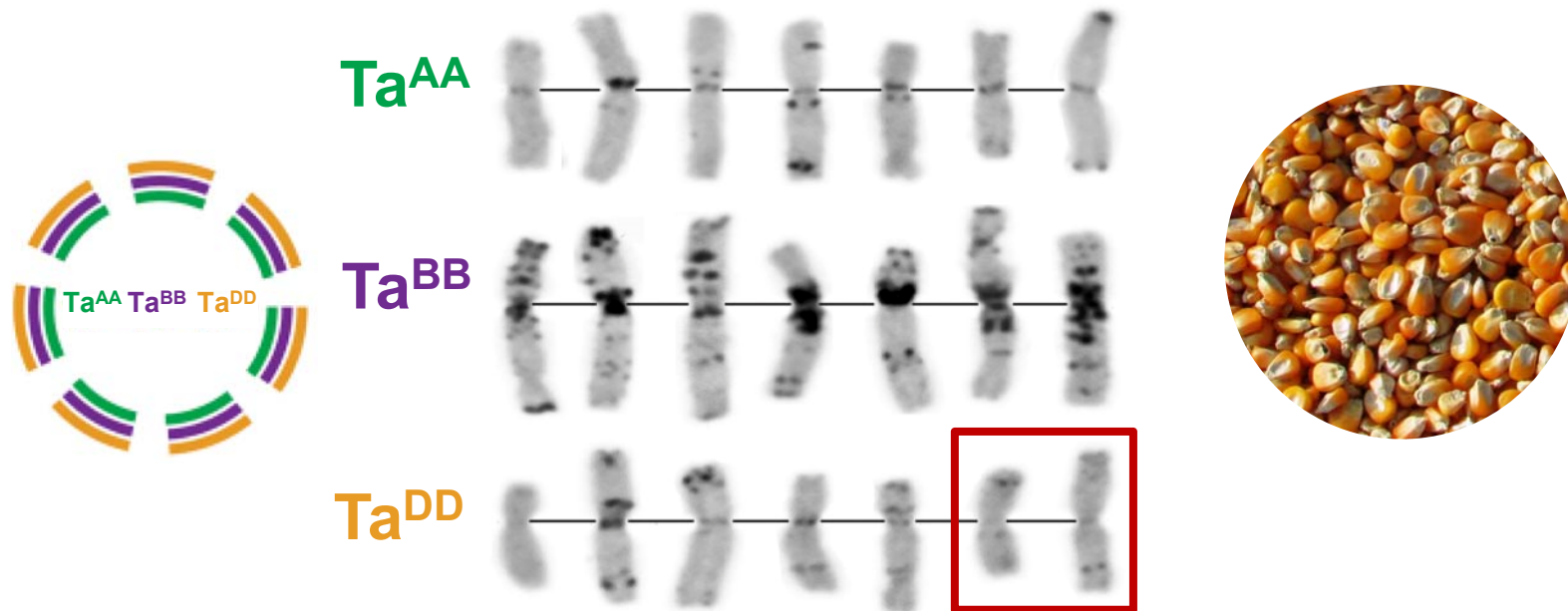
# The genome of bread wheat (*T. aestivum*) is ...



...large and complex!

- allohexaploid: three “diploid-like” ( $2n = 6x = 42$ )
- ~**17Gb** genome size; > **80% repeat** content
- Highly **similar genomes** (>97% identity)

# The genome of bread wheat (*T. aestivum*) is ...

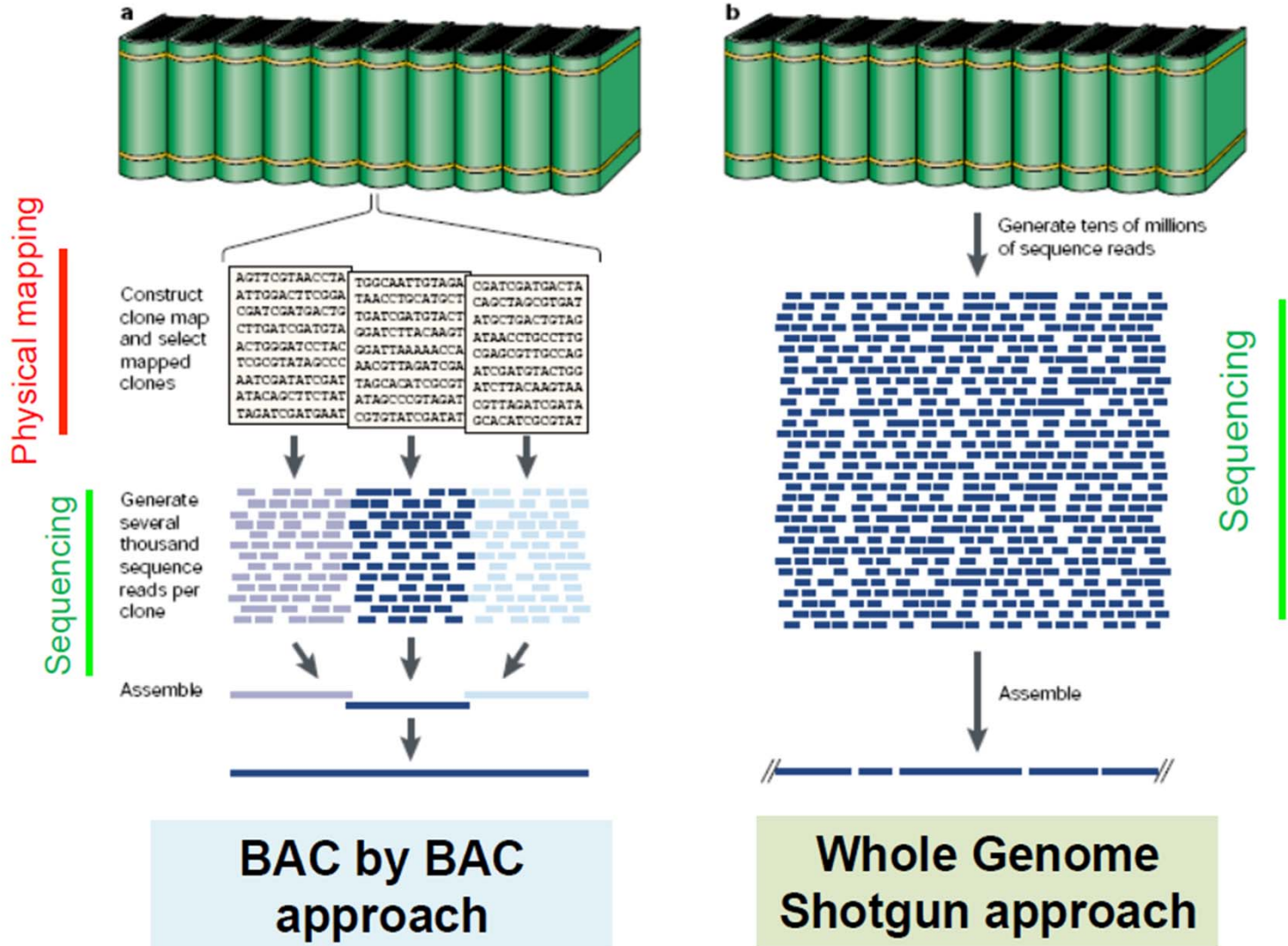


...large and complex!

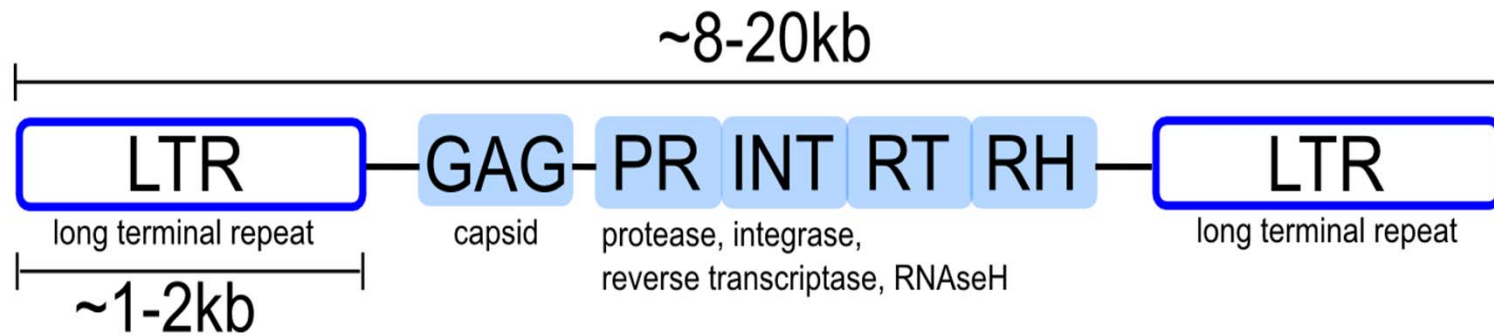
- allohexaploid: three “diploid-like” ( $2n = 6x = 42$ )
- ~17Gb genome size; > 80% repeat content
- Highly similar genomes (>97% identity)



# Sequencing complex cereal genome - strategies



# Why is the assembly so complicated?



# Wheat genome sequencing efforts



- **2014:** Chromosome-sorted WGS assembly
- ~100,000 gene models identified (highly similar A,B,D homeologous)
- BUT: highly fragmented, little inter-genic sequence available
- Newer assembly by:

**TGAC**   
The Genome Analysis Centre™

# The International Wheat Genome Sequencing Consortium



# 2015

COUNTRIES  
55



MEMBERS  
1100



INSTITUTES/COMPANIES  
372



SPONSORS  
21

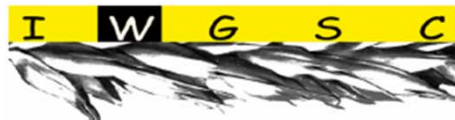


**Board of Directors**  
Role: Overall strategy and organization  
Up to 10 directors including at least 1 Outside Director

**Leadership Team**  
Role: Daily management  
Members: Appels, Eversole, Feuillet, Keller, & Rogers

**Coordinating Committee**  
Role: Scientific strategy  
Members: Sponsors & Leaders of IWGSC Projects & Initiatives

**Members**  
Role: Participation & Input  
Members: Open to Anyone



[www.wheatgenome.org](http://www.wheatgenome.org)



# Sequence Assembly Breakthrough!

**NRGene**

De**Novo**MAGIC™ 2.0

- **Novel algorithm for (genome) sequence assembly**
  - Uses standard **Illumina short reads** (at high coverage)
  - Some specific insert library sizes needed
  - **Long scaffolds** even in presence of high repetitivity and large genome sizes
  - **Fast:** 14 days for hexaploid wheat genome (17 Gb!)
-

The tragic ripples of  
an epic fraud *p. 636*

Insect pest profits from  
maize defenses *pp. 642 & 694*

Photoredox activation  
of methane *pp. 647 & 668*

# Science

\$15  
17 AUGUST 2018  
sciencemag.org

AAAS



## ROAD MAP FOR WHEAT

Ordered sequence will  
speed research *pp. 635, 661, & 662*

- Pseudo-chromosome assembly
- ~108,000 HC gene models
- ...still with gaps and missing sequence

The tragic ripples of  
an epic fraud *p. 636*

Insect pest profits from  
maize defenses *pp. 642 & 694*

Photoredox activation  
of methane *pp. 647 & 668*

# Science

\$15  
17 AUGUST 2018  
sciencemag.org

AAAS



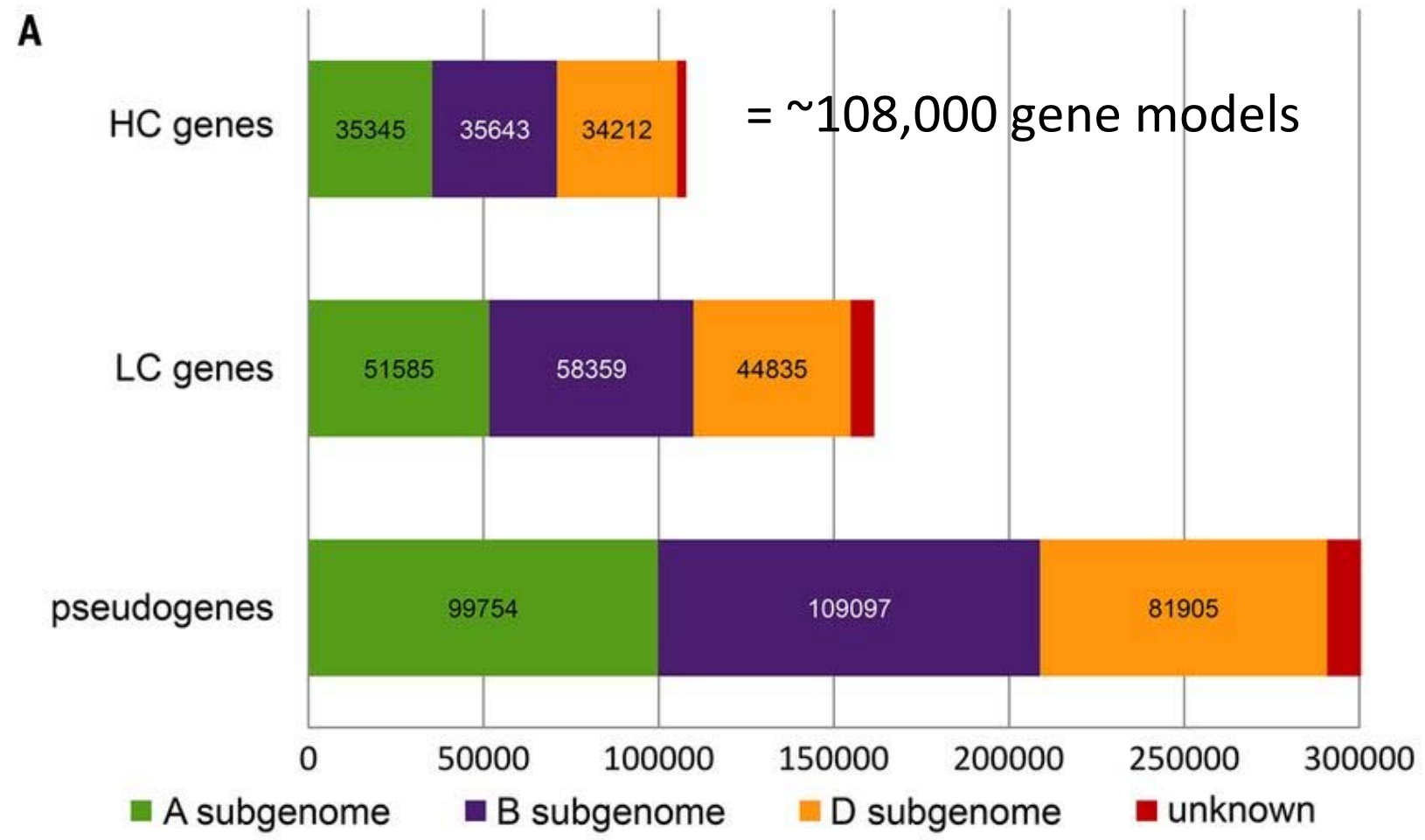
## ROAD MAP FOR **WHEAT**

Ordered sequence will  
speed research *pp. 635, 661, & 662*

**Finally!!**

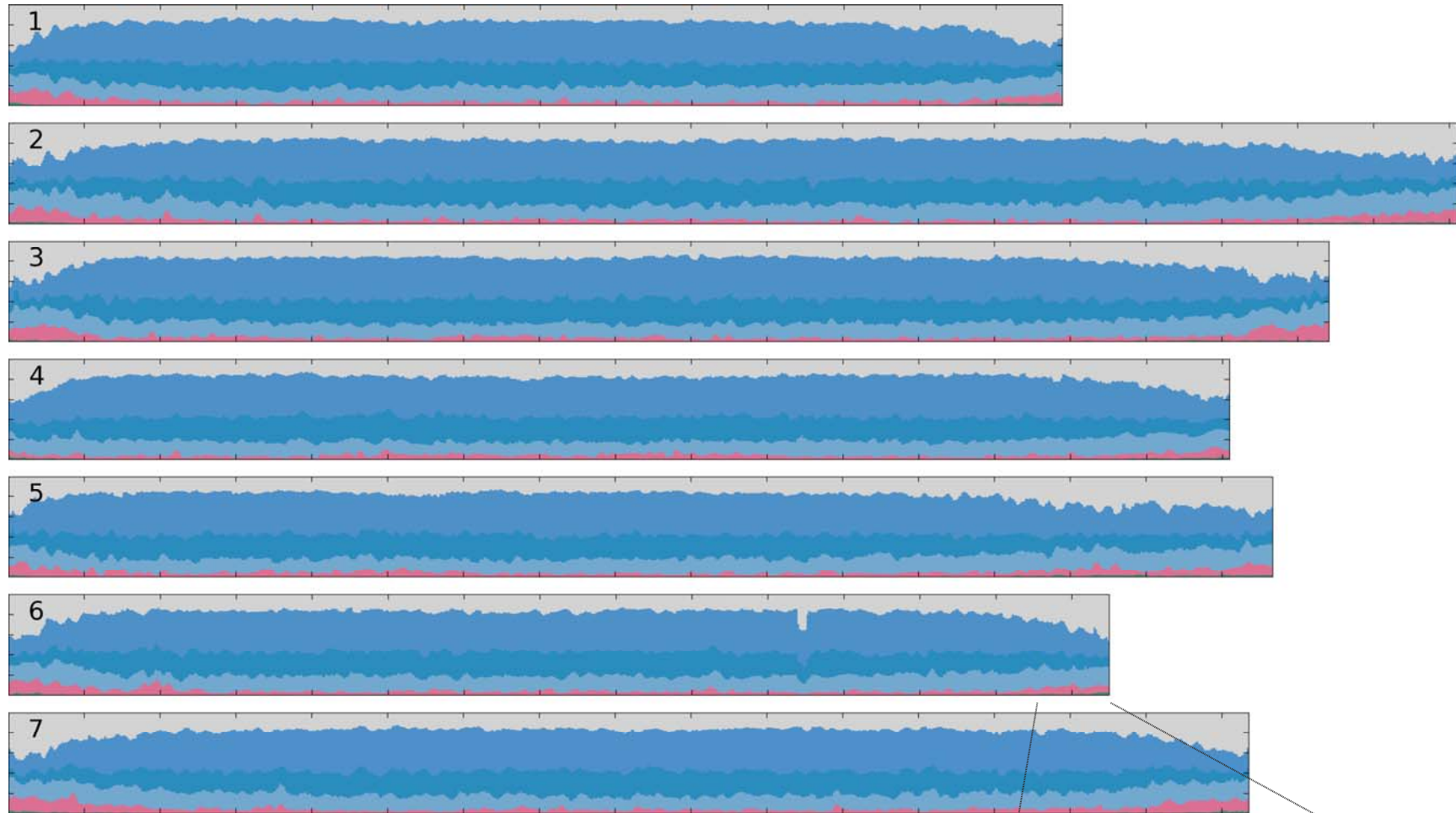
**A true reference  
genome sequence for  
bread wheat**

# The genes of bread wheat





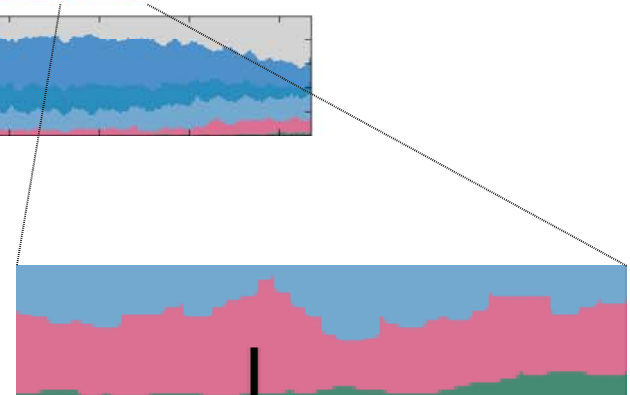
# Genes = "needles in the haystack"



## LTR-retrotransposons

- unassigned LTR-retrotransposons
- Gypsy LTR-retrotransposons
- Copia LTR-retrotransposons

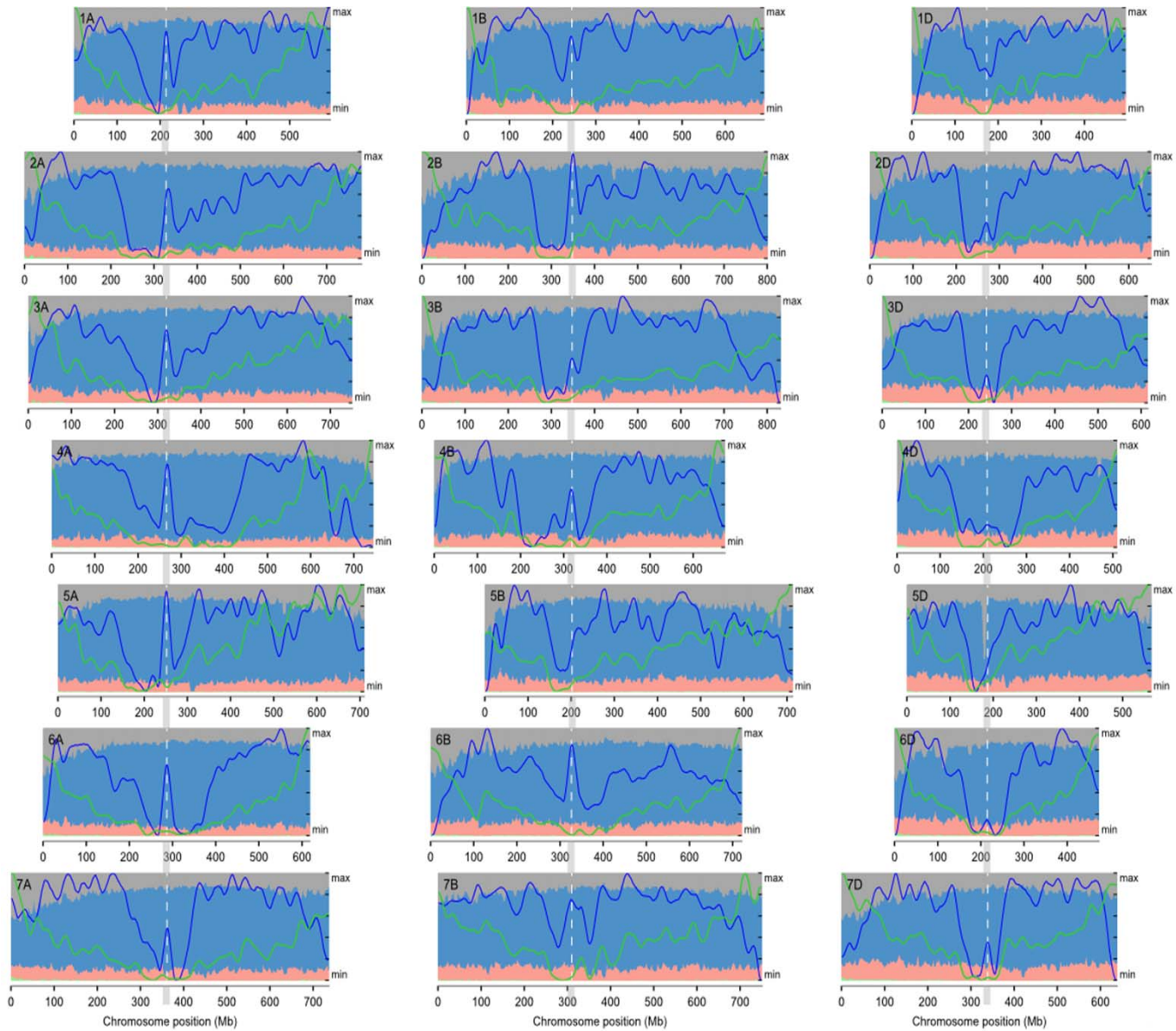
- DNA-transposons
- Genes (cds)

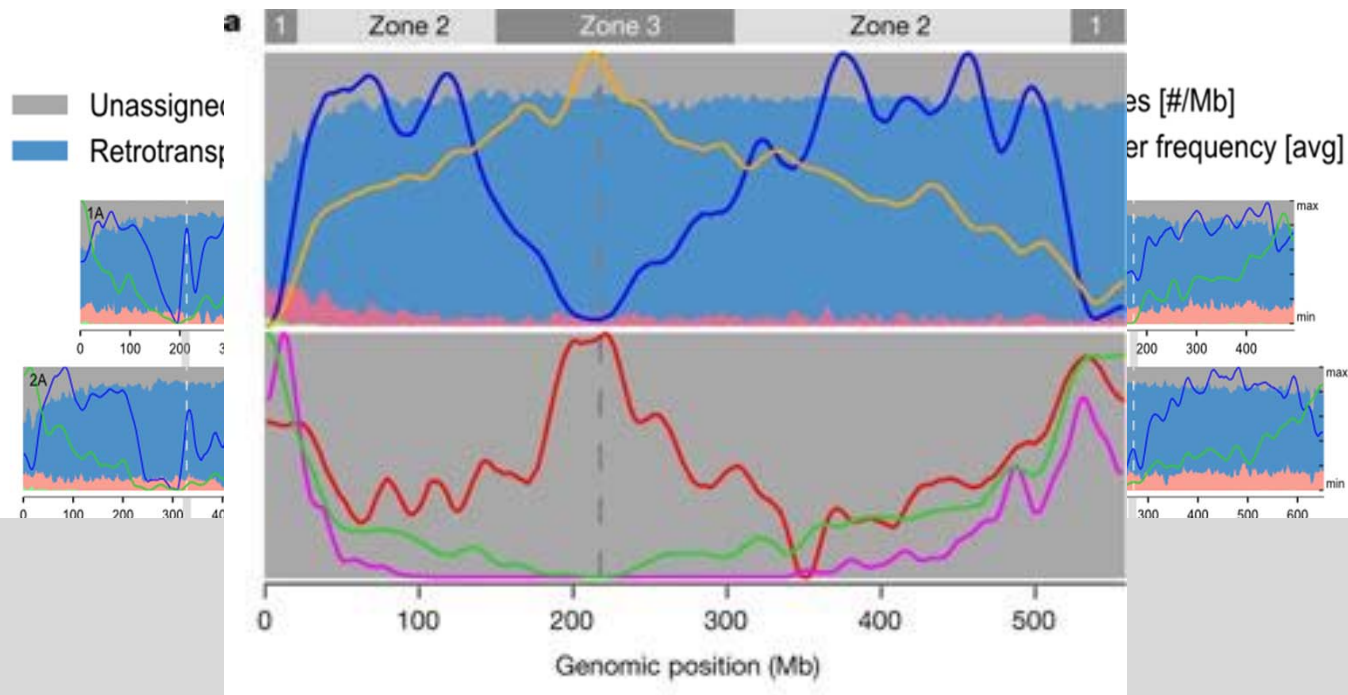


Unassigned  
Retrotransposons

DNA transposons  
Genes (CDS)

Genes [#Mb]  
20mer frequency [avg]





## Distinct chromosomal compartments:

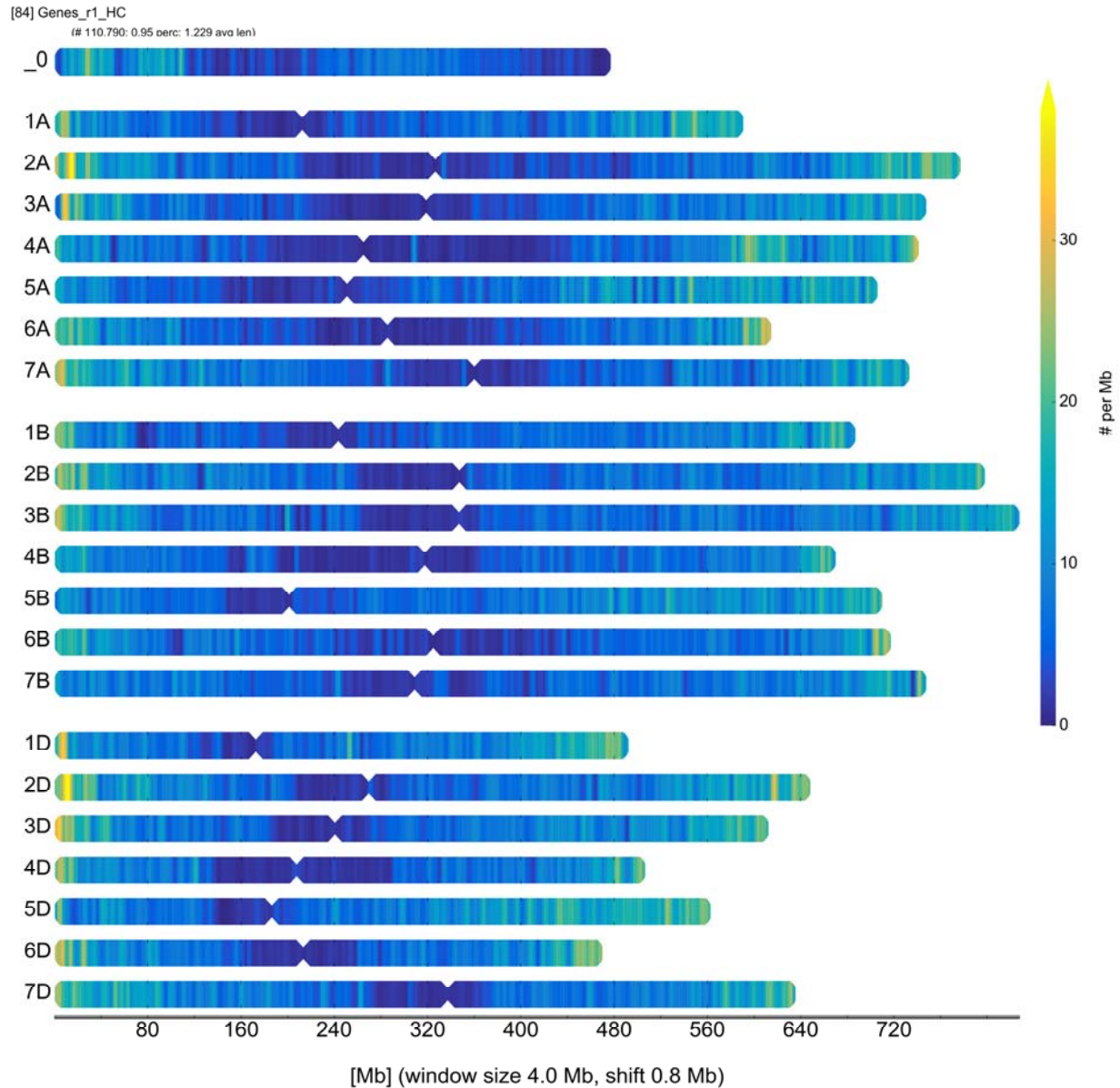
**Zone 1:** the „quite“ neighborhood:  
centromer/housekeeping

**Zone 2:** the „industrial area“: middle parts

**Zone 3:** the „innovation quarter“: chromosome  
ends/high recombination



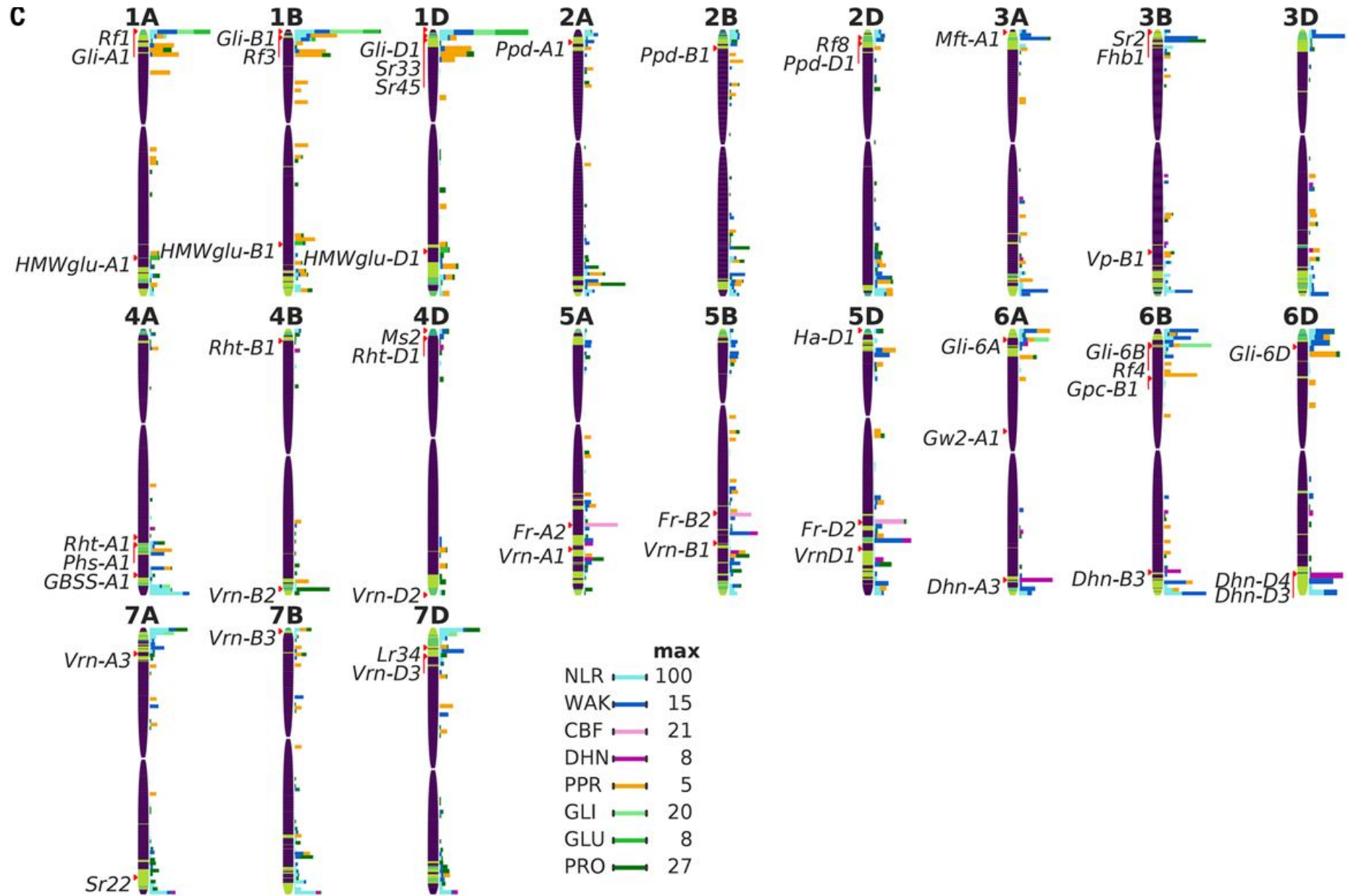
# Genomic distribution of genes



**What is it good for?**



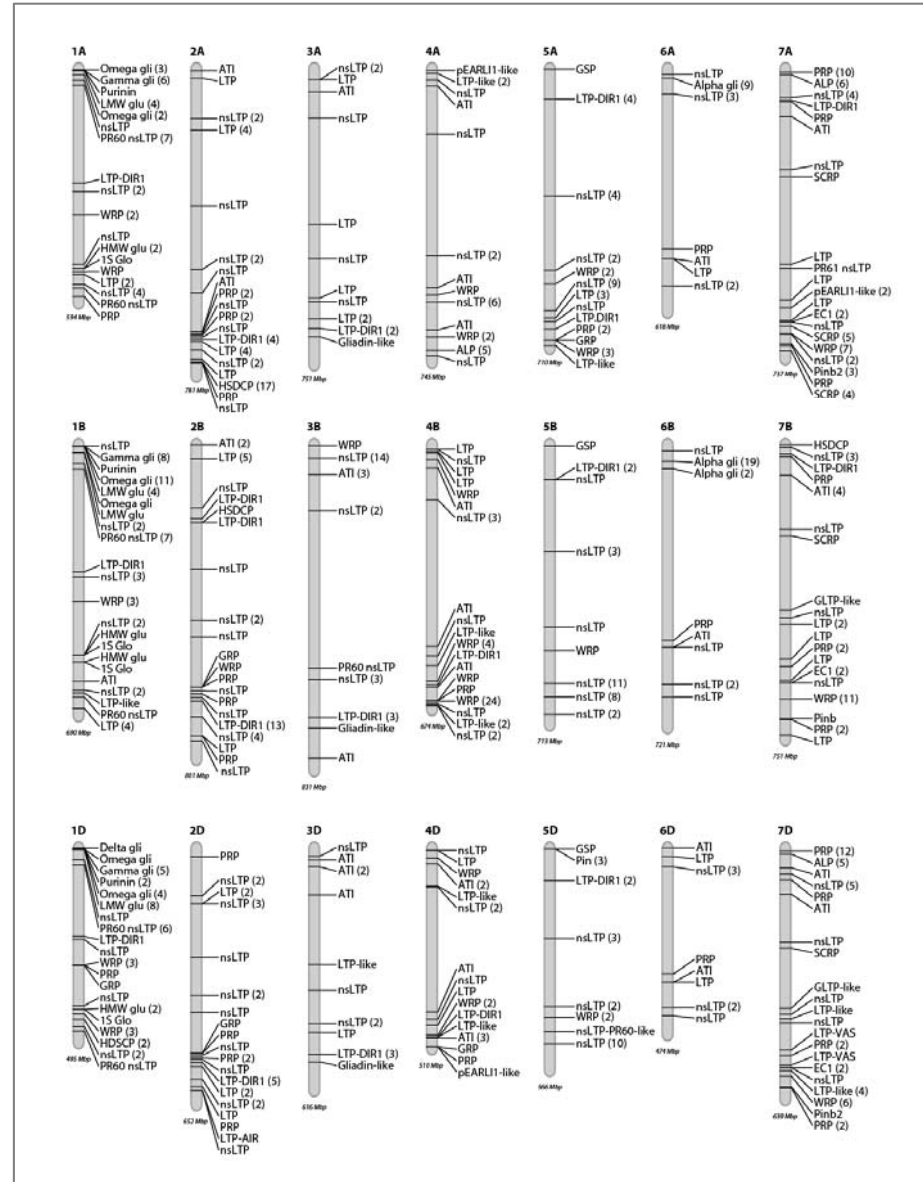




# First comprehensive map of wheat prolamins

IWGSC, Science 2018

- Within the IWGSC RefSeq v1 annotation, 731 proteins were manually corrected, including 135 proteins that were added as a completely new sequence
- Expressed everywhere e.g. in roots, leaves, spike, pollen or grain



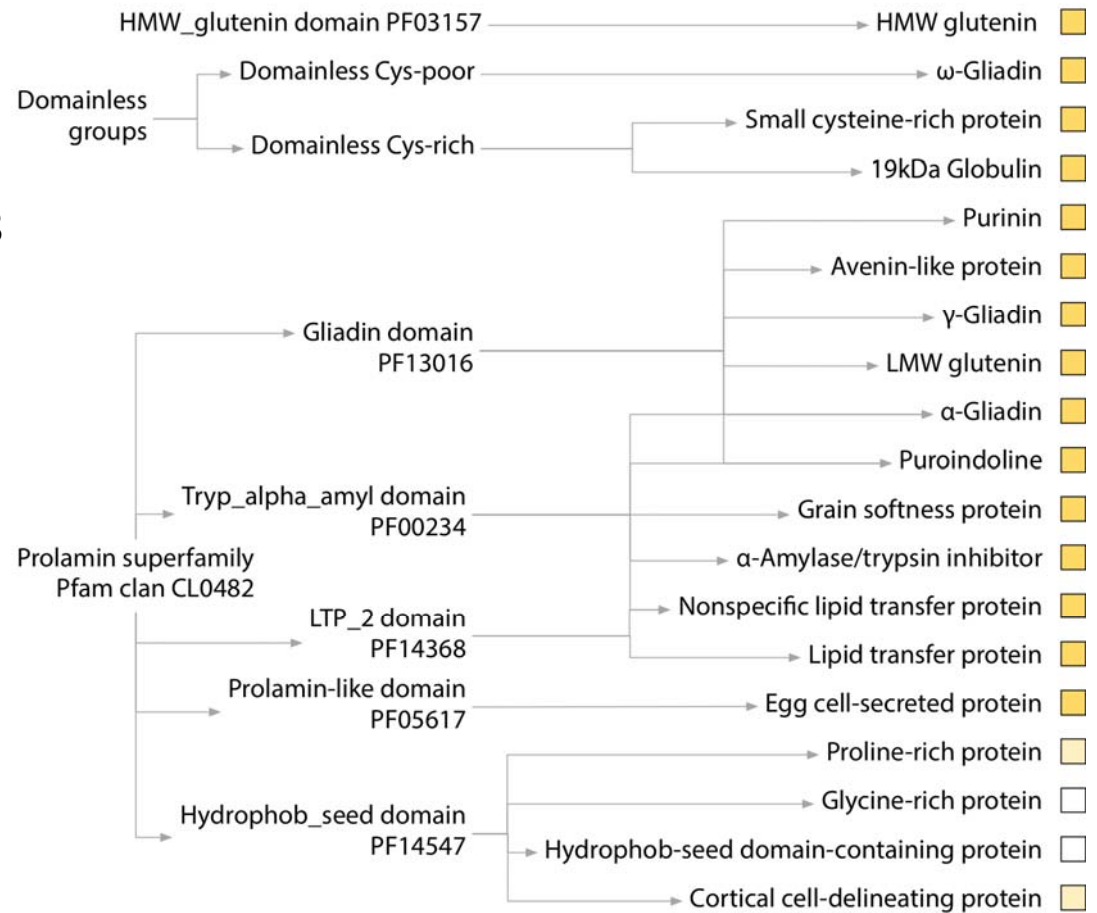
# The wheat prolamins superfamily

Juhasz et al., Science Advances, 2018

Pfam domains – Prolamin clan CL0482

- Gliadin
- Tryp-alpha-amyl
- LTP2
- Hydrophob\_seed
- Prolamin-like

- HMW glutenins
- Domainless protein groups
- omega gliadins
- small cysteine-rich proteins

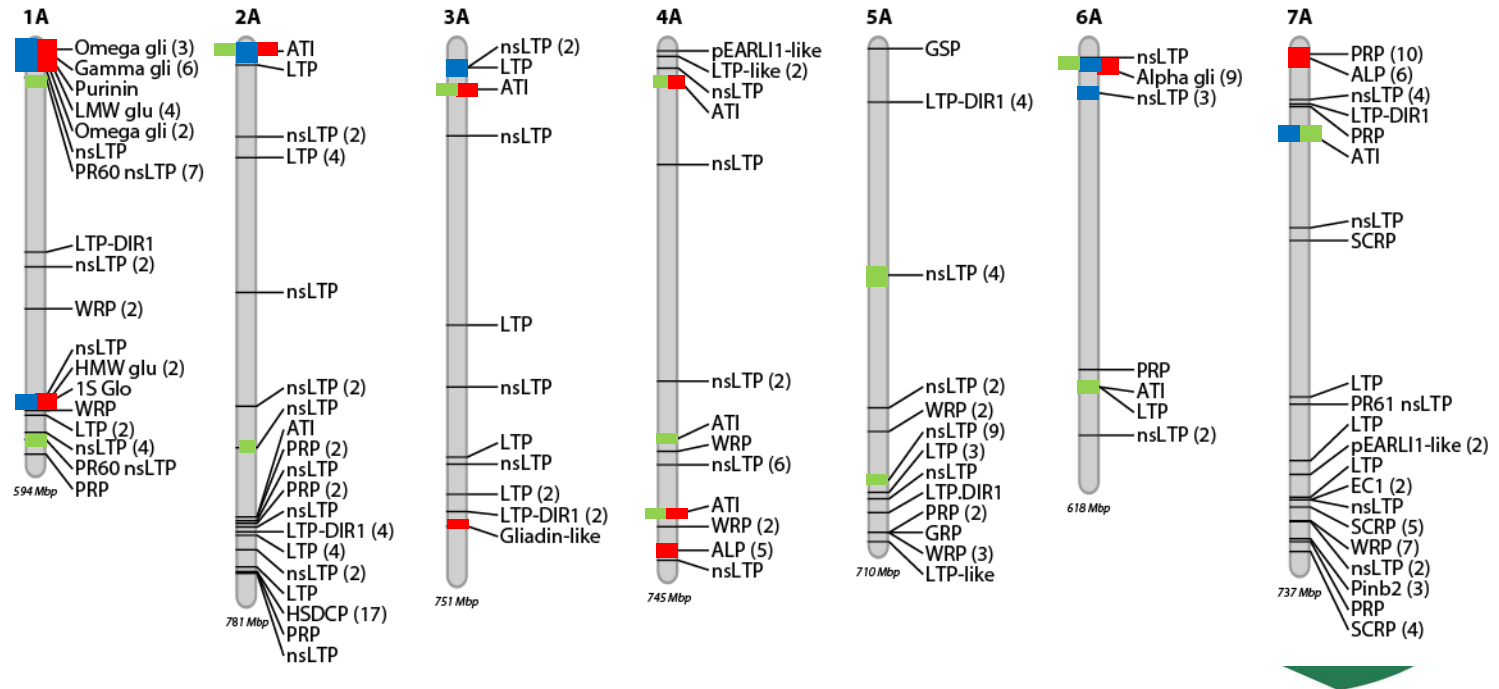


# The classic prolamins as immune reactive proteins

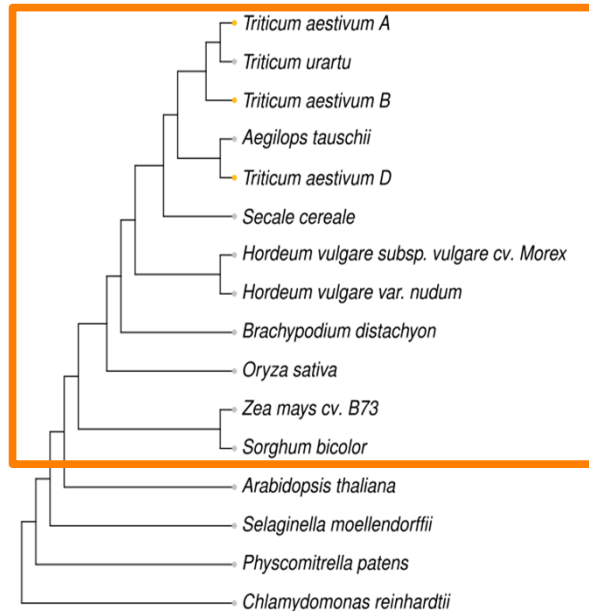
Juhasz et al., Science Advances, 2018

## A genome

- Celiac disease
- Wheat allergy
- Baker's asthma



# The Prolamin Clan in Grasses



## Creation of gene family dataset:

- Protein sequences screened for known domains using HMMER3
- protein sequences were clustered by similarity NCBI blast+
- “gene families” were constructed in the form of orthologous groups using OrthoFinder
- All protein sequences with a HMW glutenin or a Prolamin clan domain were extracted → **1783 sequences**

Juhász et al. 2018, Sci. Adv.

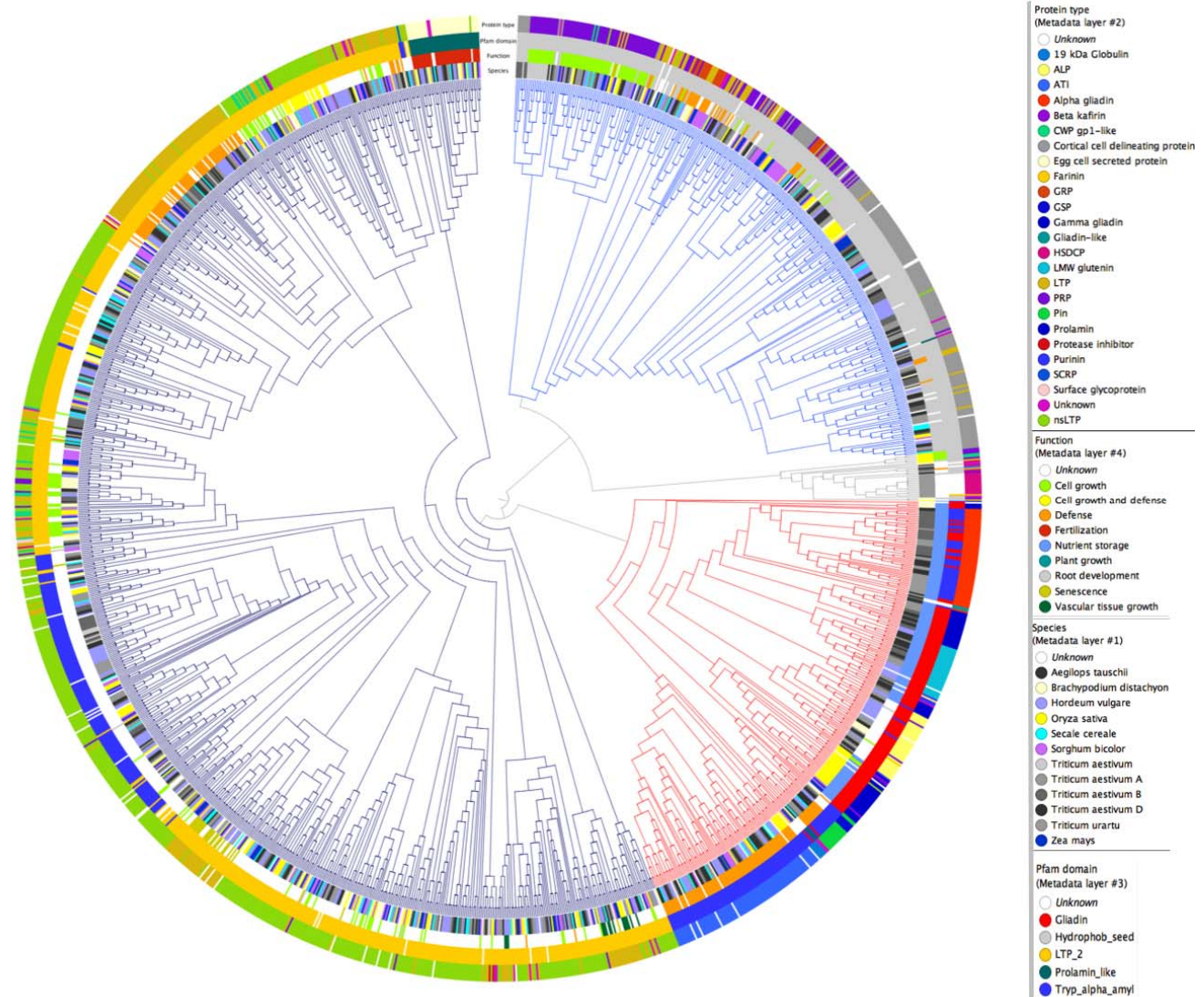


# Tree of Prolamin Clan sequences

Hydrophob seed:  
Cell growth, Root  
development, defense

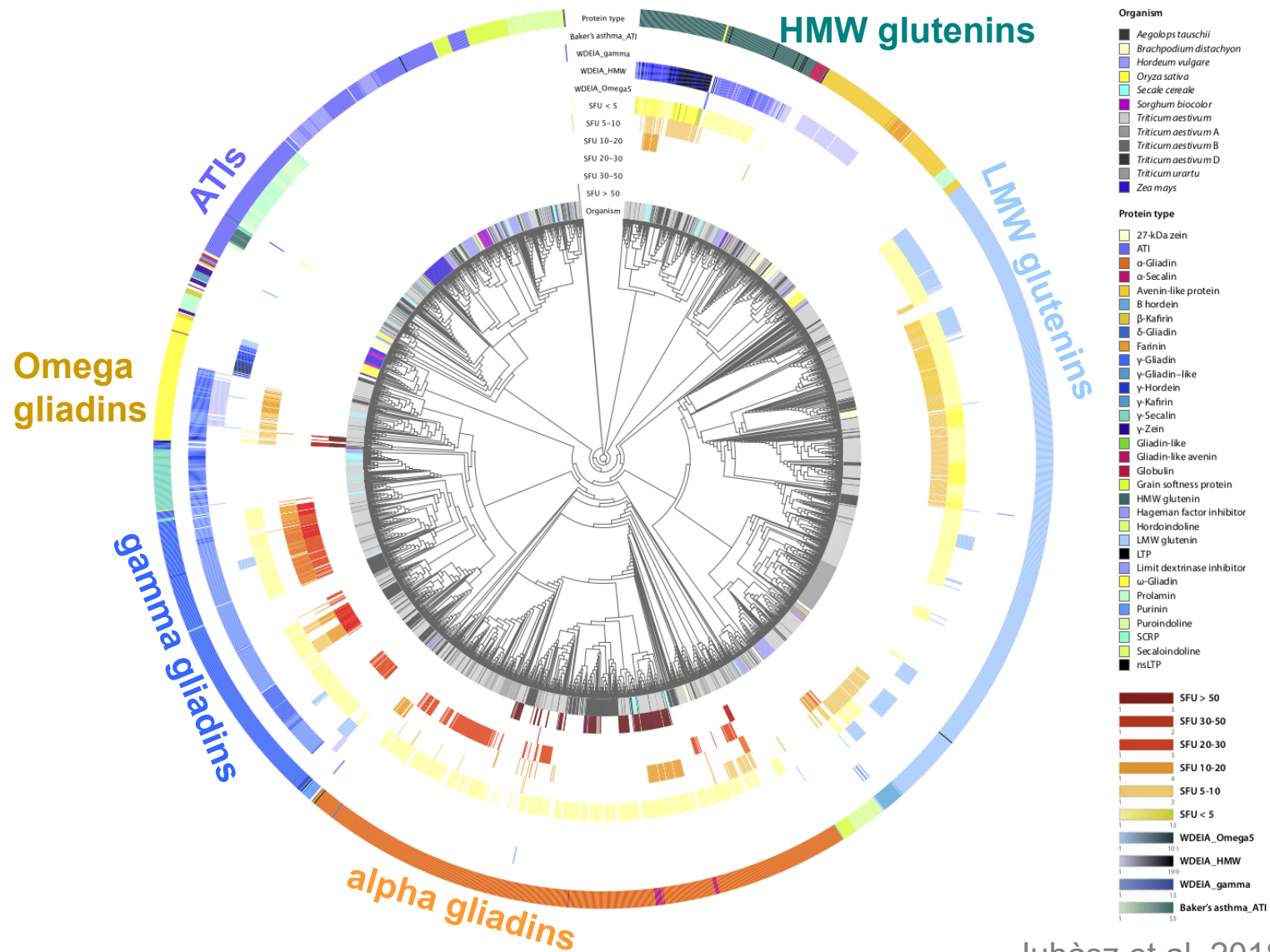
LTP\_2,  
Tryp\_alpha\_amy:  
Cell growth,  
Senescence  
Prolamin-like :  
Fertilization

Gliadin,  
Tryp\_alpha\_amy:  
Nutrient storage,  
defense



Juhász et al. 2018, Sci. Adv.

# Celiac disease & Baker's asthma epitopes



Juhász et al. 2018, Sci. Adv.

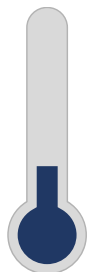
# Seed allergen expression in different environments

Juhasz et al., Science Advances, 2018



## High temperature

Increasing effect on the celiac disease associated gene expressions – gliadins and glutenins

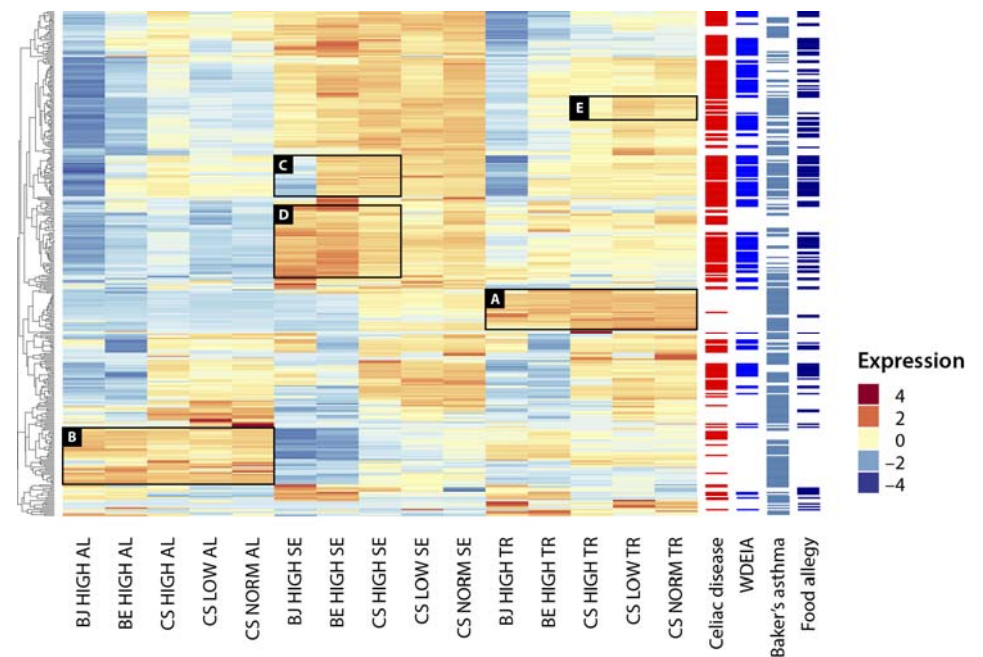


## Low temperature

Increasing effect on baker's asthma and food allergy genes

Chinese Spring  
Bjarne  
Berserk

Low: 15/10°C day/night  
Normal: 20/16°C day/night  
High: 26/20°C day/night



**What's next??**

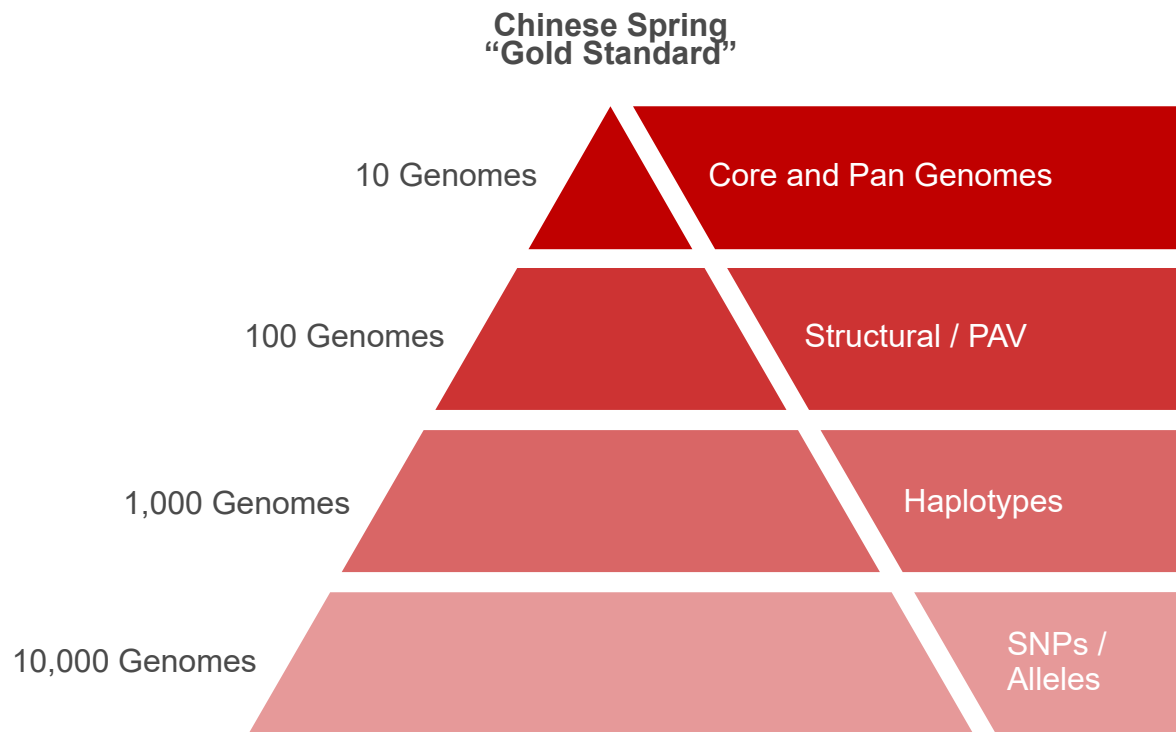




**One reference genome is not enough!**







# The Ten+ Genomes Project

NRGene  
RefSeq  
Assemblies

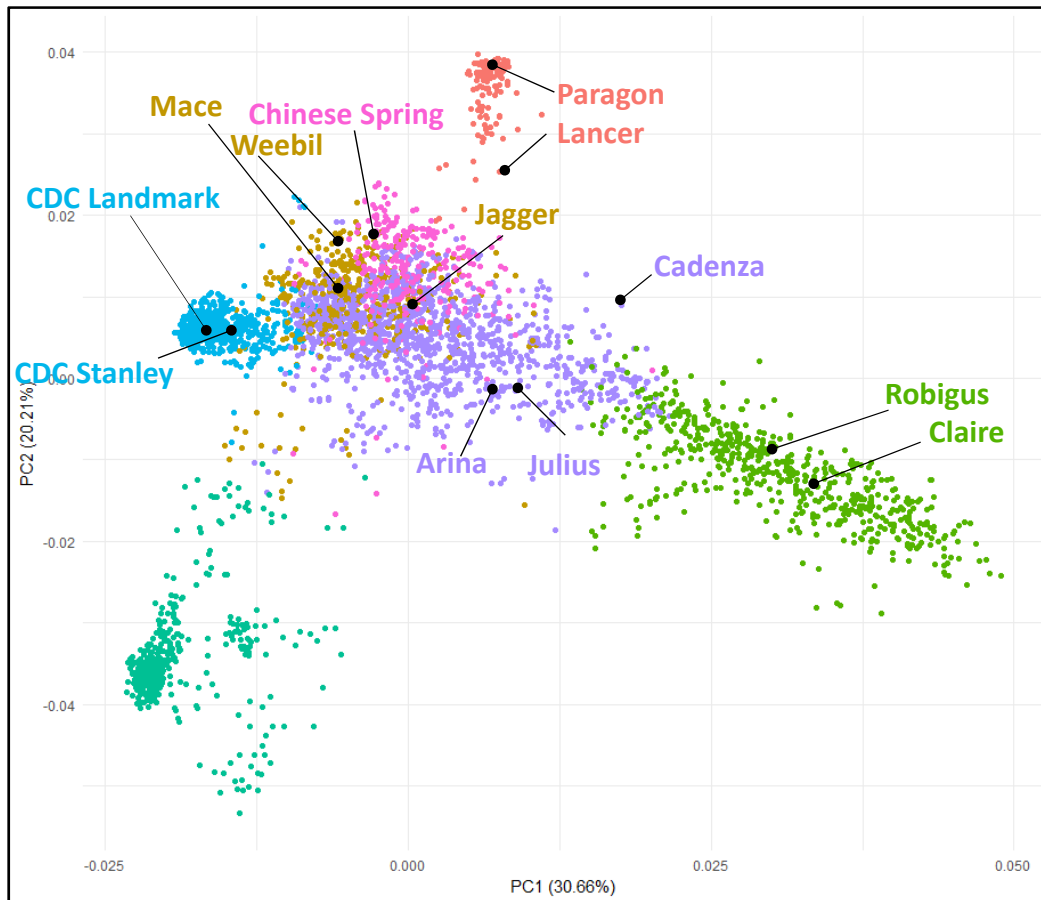


- 2 Canadian spring wheat (**CDC Landmark, CDC Stanley**)
- 1 USA (**Jagger**) - J. Poland, G. Muehlbauer
- 1 German winter wheat variety (**Julius**) – N. Stein, K. Mayer
- 1 Swiss winter wheat variety (**Arina**) – B. Keller
- 2 Australian varieties (**Mace, Lancer**) – P. Langridge
- 1 Japanese variety (**Norin61**) – K. Shimizu
- 2 Chinese varieties (**Kenong 9204\***) HongQing Ling
- Syngenta – (**SY Mattis**)

W2RAP  
Assemblies

Cadenza, Paragon, Kronos, Robigus, Claire

# Harnessing the diversity from global breeding programs



## Genomic Diversity Analysis:

- 3,819 samples (mostly hexaploid)
- 1,779 SNP markers from the 35K and 90K arrays, distributed across all chromosomes
- PCA transformed and clustered using k-means

# Tetraploid wheats – the wild and the domesticated



**wild emmer wheat**

*T. turgidum* ssp. *dicoccoides*



**Durum (pasta) wheat**

*T. turgidum* ssp. *durum*

- A and B subgenome, genome size ~10 Gb
- plus diploid wheats and ancestors + Rye & Spelt

# Perspectives...what else is it good for?

- ✓ Gene and marker discovery!
  - ✓ Predictive Breeding/genomic selection
  - ✓ Accessing variation
    - ✓ Reduced genetic variation in modern breeding lines;
    - ✓ Only 10% of genetic variation in wheat has been captured in modern varieties
  - ✓ Comparative Triticeae Genomics – the non-recombining part of the genome, gene regulation/networks
  - ✓ Towards a less allergenic wheat...
-



# Meet the team...



## PGSB

Daniel Lang  
Heidrun Gundlach  
Thomas Lux  
Iris Fischer  
Michael Seidel  
Verena Prade  
Nadia Kamal  
Georg Haberer  
Sven Twardziok  
Klaus Mayer

## Murdoch University

Angela Juhasz  
Rudi Appels

## Norwegian Academy of Lifesciences

Tetiana Belova  
Odd-Arne Olsen



[www.wheatgenome.org](http://www.wheatgenome.org)



# Acknowledgements Wheat

## IWGSC Sponsors



I W G S C



# Acknowledgements Wheat



**IPK**  
Nils Stein  
Martin Mascher

**IPK BIT**  
Sebastian Beier  
Uwe Scholz

**IEB Olomouc**  
Hana Simkova  
Jaroslav Dolezel

**INRA**  
Fred Choulet  
Etienne Paux  
Michael Alaux  
et. al.

**PGSB**  
Heidrun Gundlach  
Thomas Lux  
Iris Fischer  
Daniel Lang  
Verena Prade  
Georg Haberer  
Michael Seidel  
Sven Twardziok  
**Klaus Mayer**

**Murdoch University**  
Angela Juhasz  
Rudi Appels

**Norwegian Academy of Lifesciences**  
Tetiana Belova  
Odd-Arne Olsen

**IWGSC: Kellye Eversole, Jane Rogers**  
**Bayar: Catherine Feuillet**  
**Univ. of Zurich: Thomas Wicker & Beat Keller**  
**Univ. of Udine: Michele Morgante et al.**  
**KWS: V Korzun**  
**CNRGV: H Berges, A Bellec**  
**Haifa: A Korol, Z Frenkel**  
**KSU: Jesse Poland**  
**TGAC: Mario Caccamo et al.**  
**USASK: Curtis Pozniak**  
**U.Tel Aviv: Assaf Distelfeld**  
**John Innes Center: Cristobal Uauy et al**  
**Earlham Institute: Anthony Hall et al.**

**WheatScan partners**  
Katharina Scherf



Bundesministerium  
für Ernährung  
und Landwirtschaft



Bundesministerium  
für Bildung  
und Forschung





**Thank you for your attention!**

---