I W G S C
International Wheat Genome
Sequencing Consortium

IWGSC Phase II Activities: Moving from Structural to Functional Genomics

IWGSC Wheat Genome Manual and
Functional Annotation Workshop
14 January 2020

PAG XXVIII

Kellye Eversole
IWGSC Executive Director

IWGSC

The International Wheat Genome Sequencing Consortium

BASF We create chemistry

FLORIMOND DESPREZ

INRAE

Limagrain from earth to life

KANSAS WHEAT

RAGT

syngenta

THE UNIVERSITY of ADELAIDE

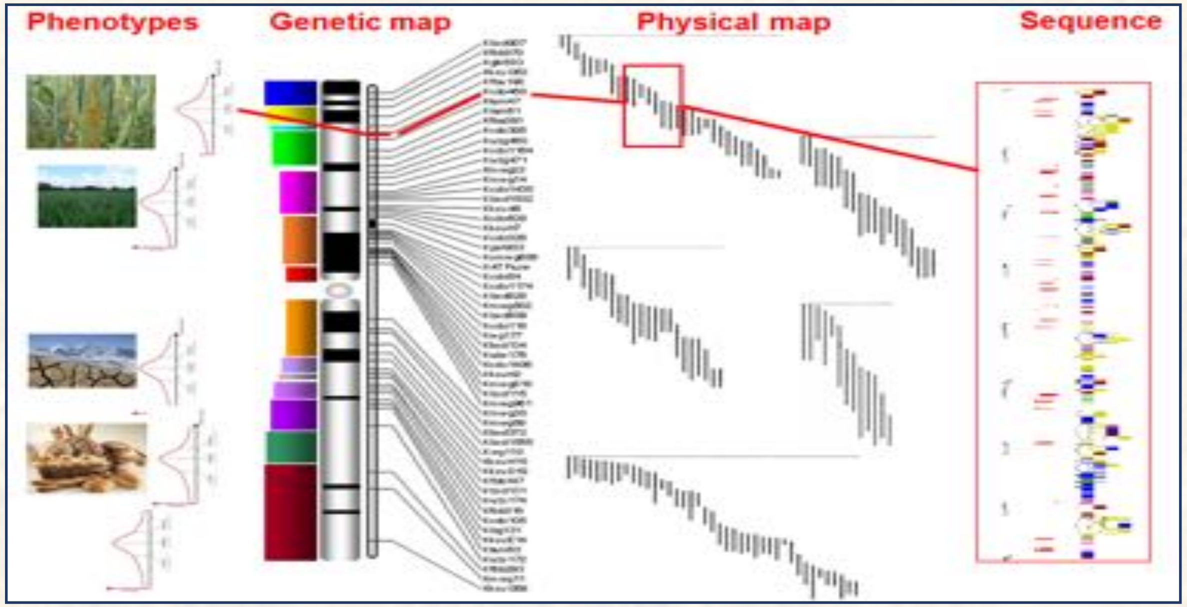
2005 – Goal & Vision

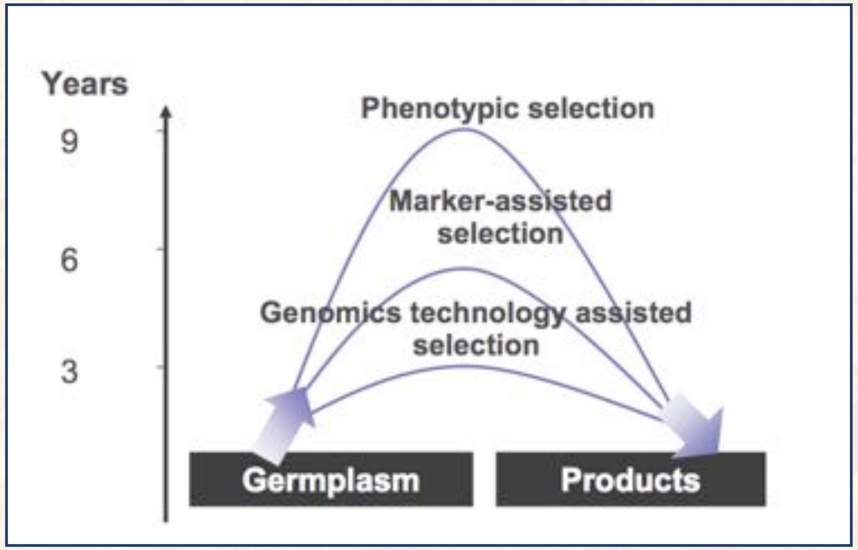
Goal

- Lay a foundation to accelerate wheat improvement
- Increase profitability throughout the industry

Vision

- High quality annotated genome sequence, comparable to rice
- Physical map-based, integrated and ordered sequence





High Quality Reference Sequence of Bread Wheat: IWGSC RefSeq v1.0

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

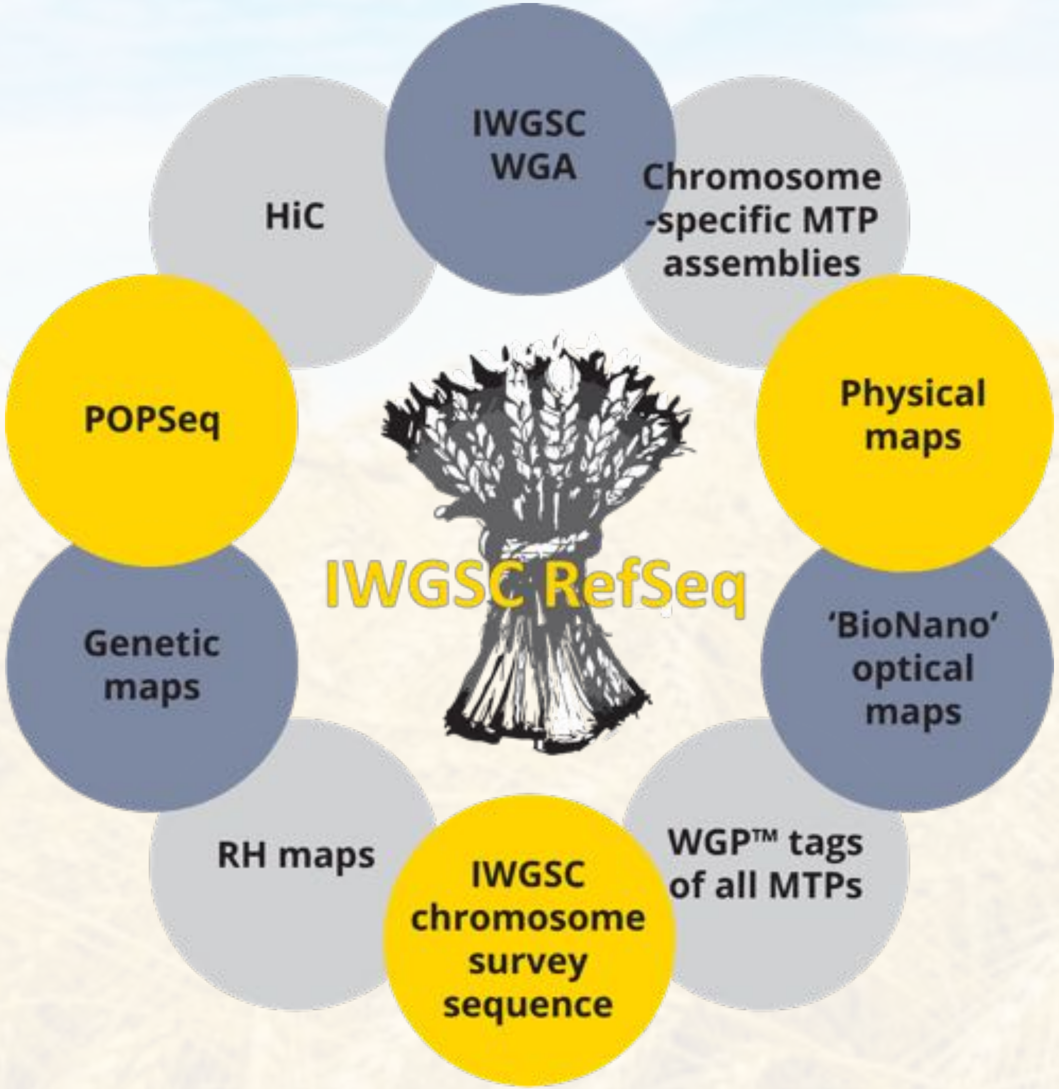
$15
17 AUGUST 2018
sciencemag.org

AAAS

ROAD MAP FOR
WHEAT

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

Science, 17 August 2018



IWGSC RefSeq

IWGSC WGA

Chromosome -specific MTP assemblies

Physical maps

'BioNano' optical maps

WGP™ tags of all MTPs

IWGSC chromosome survey sequence

RH maps

Genetic maps

POPSeq

HiC

Measuring the Impact of IWGSC RefSeq v1.0

> 630 papers*

2017-2019

Genome structure (89)

Genome expression (44)

Genetic diversity (38)

Trait genomics (337)

Genome enabled breeding (81)

* Incomplete number of papers published that referenced or used IWGSC RefSeq and associated resources

Vision and strategy for IWGSC 2.0

IWGSC

International Wheat Genome Sequencing Consortium

Reference sequences with high quality functional annotations

Public, user-friendly, integrated databases and tools for all users

A wheat diversity panel representing the breadth of global wheat diversity

An international, well connected and coordinated community

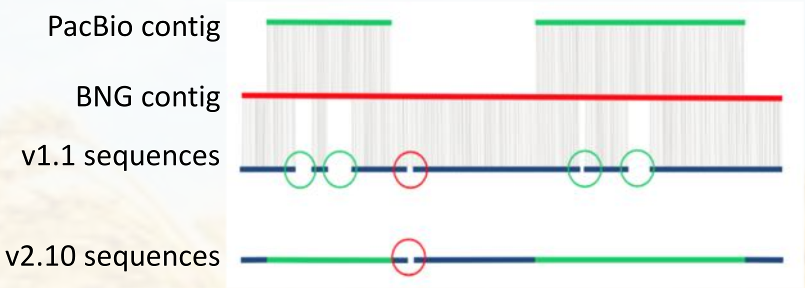
IWGSC 2.0 Vision

Enhance breeding through an increased understanding of the molecular basis of traits and their allelic diversity

IWGSC RefSeq v2.0

Under the leadership of Mingcheng Luo and Jan Dvorak (UC Davis, CA, USA),
with funding from the US National Science Foundation

Gap closing with PacBio contigs



Number of gaps:

v1.0: 527,170 → v2.0: 183,604

↓ 65%

Improvements v2.0 vs v1.0:

- Resolved 75 chimeric scaffolds
- Anchored 283 additional scaffolds from ChrUn of v1.0
- Corrected orientation errors for 357 scaffolds
- Revised locations (change of order) for 244 scaffolds
- Change of total length of pseudomolecules: + 158,997,657 bp
- Change of total effective length of pseudomolecules: + 167,843,761 bp
- Percentage of sequences anchored on chromosomes: 97.65% in v2.0 (96.69% in v1.0)
- All revisions: 1,380 Mb (≈ 10%) of the v1.0 assembly

Caution:

- **Between 5-10% of the genes in v2.0 may be incorrect**
- **Some correct genes in v1.0 became pseudogenes in v2.0**
- **A list of all of the changed regions is being prepared and will be released shortly**
- **Mingcheng Luo and Fred Choulet are working to resolve all errors in v2.0**

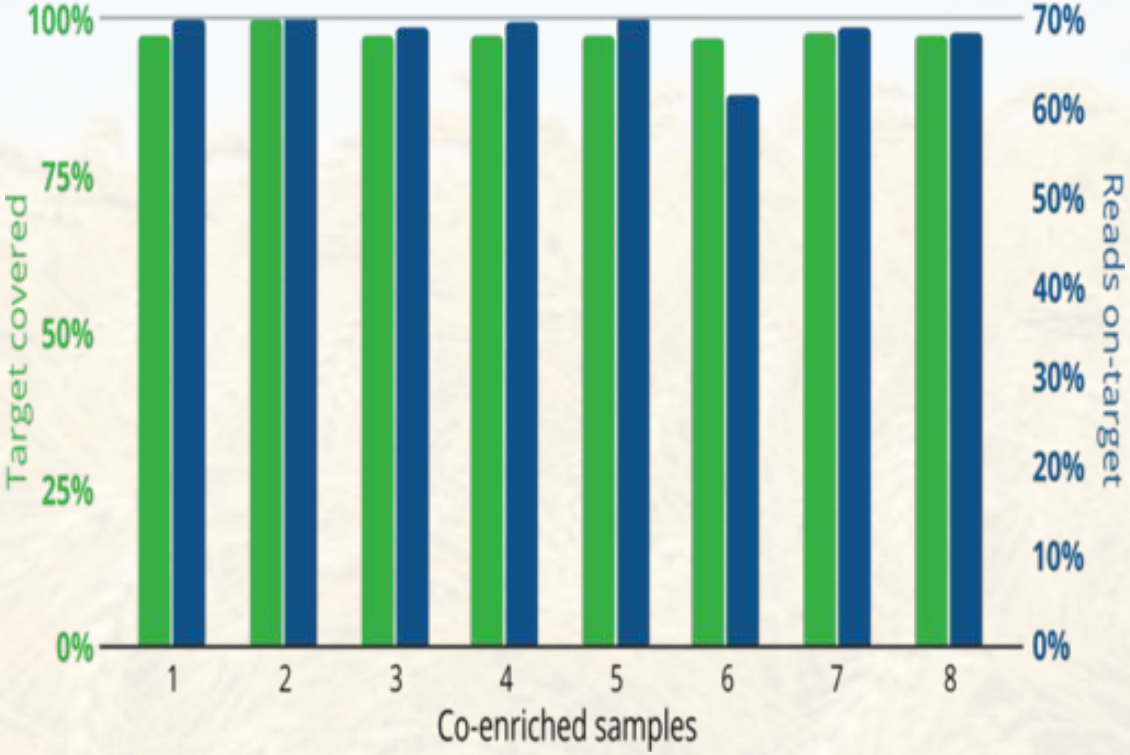
IWGSC-Arbor Biosciences Collaboration

IWGSC Collaboration with Arbor Biosciences

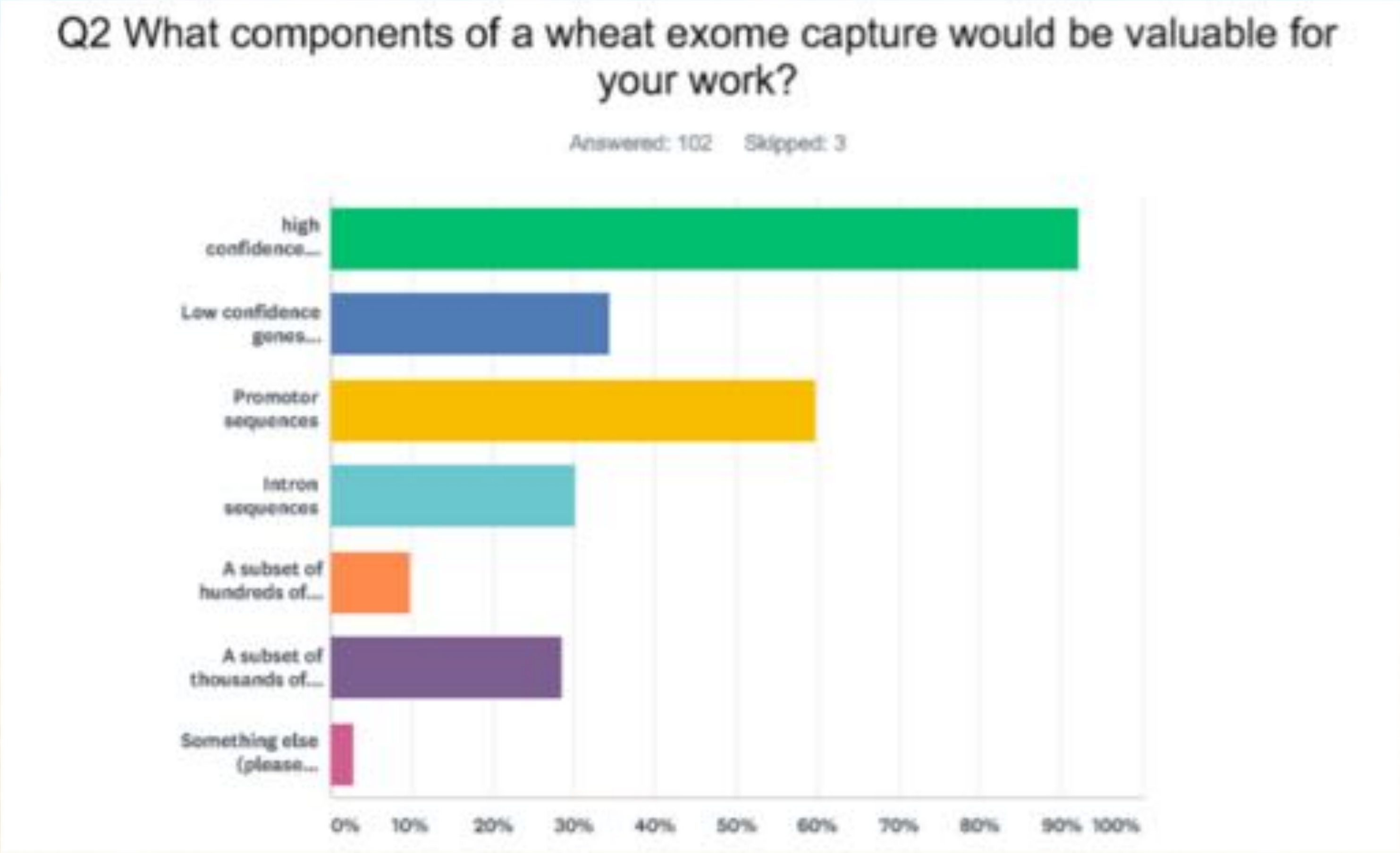
- Based on Chinese Spring (Hexaploid)
- Compatible with other cultivars
- Focused on High Confidence Exons
- 200 Mb of 17 Gbp Genome
- 30x Coverage with 18 Gb Sequencing

myBaits Wheat Exome v1

- Various Kit Sizes
    - 8, 48, 96 reaction kit sizes
    - Reaction sufficient for 8 pooled samples
- Services offered through myReads™
- New Modules Coming Soon
    - Low-confidence annotations
    - Promoter regions



IWGSC-Arbor Biosciences Expanded Collaboration


Q2 What components of a wheat exome capture would be valuable for your work?

Answered: 102 Skipped: 3

Promoter Capture Project Leaders


Jorge Dubcovsky


Etienne Paux

Promoter capture composition

Development is accelerated

Putative promoter regions

+ Existing capture performance data

+ ATAC-seq-identified regions

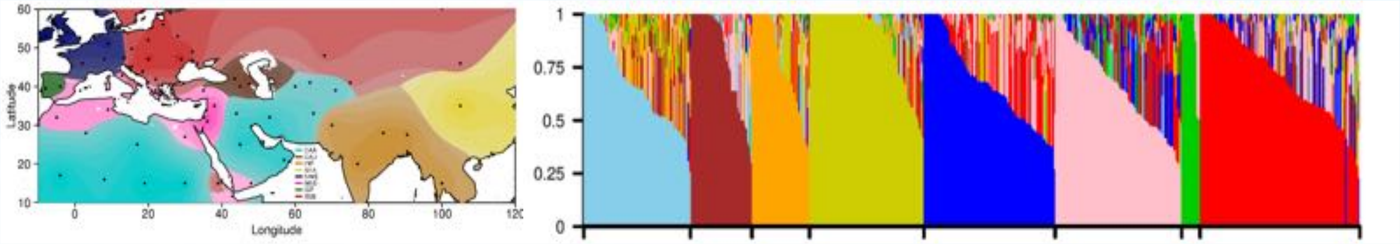
- kmer-based repeat filtration

180 Mbp

Synthesis starts in February

Slide provided by Jacob Enk

Towards an IWGSC Wheat Diversity Panel



Platinum quality sequences and assemblies of Eight landraces with less than 1% admixture

➔ describe ancient wheat haplotypes

+

Additional lines (including elite varieties) @ 30-40X

➔ describe the modern genetic diversity (including alien introgressions)

Courtesy of Etienne Paux

Balfourier, et al. https://doi.org/10.1126/sciadv.aav0536

IWGSC Manual & Functional Annotation

- Release of annotation v2.0

- Exploration of artificial intelligence (AI) and machine learning (ML) tools for functional annotation of the wheat genome
  - ✓ IWGSC-BASF project will be developed to explore deep learning (DL) as a means of accelerating functional annotation of IWGSC RefSeq (Xi Wang & Kellye Eversole leading)
  - ✓ Working with Tanya Berardini to use ML (natural language processing – NLP) to automate literature curation for wheat gene function

IWGSC Pre-Publication Access to Additional Sequences & Data

- We will be releasing the sequence of an elite Turkish variety and exome data in June 2020

- Encourage all scientists to make genome sequences and data available under the Toronto protocol through the IWGSC repository at URGI

IWGSC RefSeq Data Access & Availability

https://wheat-urgi.versailles.inra.fr/Seq-Repository/

All data available at URGI

- IWGSC RefSeq assembly v1.0
- IWGSC RefSeq assembly v2.0 (under Toronto protocol)
- IWGSC RefSeq annotation v1.1 including genes and RNAseq mapping
- IWGSC RefSeq annotation v1.0 including structural annotation (genes, transposable elements, ncRNAs), functional annotation, RH maps, GBS maps, optical maps)
- Physical maps for all chromosomes
- MTP BAC WGP^{TM} sequence tags for all chromosomes, except 3B

Non-Toronto Protocol Data also available at Ensembl Plants, NCBI and GrainGenes

IWGSC Webinars

- 22 January: Understanding abiotic stress signaling in wheat through (phospho)proteomics

- Register: https://t.co/4VVwN5npNa?amp=1
- Soliciting ideas for additional IWGSC Webinars on technologies or research results

IWGSC Sponsors: Thank you!

BASF
We create chemistry

FLORIMOND
DESPREZ

INRAE

Limagrain
from earth to life

KANSAS
WHEAT

R·A·G·T

syngenta

THE UNIVERSITY
of ADELAIDE

I W G S C

Thank you for your attention!

www.wheatgenome.org

f wheat.genome

@wheatgenome