IWGSC 2.0: Wheat Genome Curation for Future Crop Resilience The International Wheat Genome Sequencing Consortium

Catalina Johnson, Joshua C Stein, Victor Llaca, Kevin Fengler, Isabelle Caugant, Elodie Belmonte, Vincent Pailler, Hugo Zalzale, Leonardo Pereyra-Bistraín, Clément Debiton, et al.

To cite this version:

Catalina Johnson, Joshua C Stein, Victor Llaca, Kevin Fengler, Isabelle Caugant, et al.. IWGSC 2.0: Wheat Genome Curation for Future Crop Resilience The International Wheat Genome Sequencing Consortium. PAG32, 2025, San Diego, United States. ⟨hal-05519247⟩

HAL Id: hal-05519247
https://hal.inrae.fr/hal-05519247v1

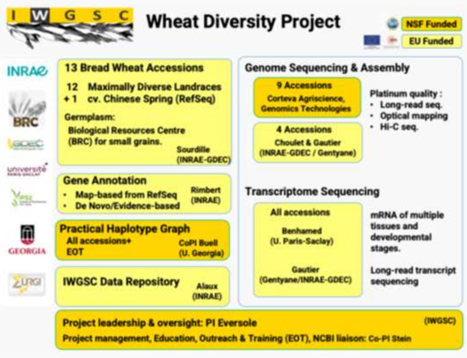
Submitted on 19 Feb 2026

IWGSC 2.0: Wheat Genome Curation for Future Crop Resilience

The International Wheat Genome Sequencing Consortium[1], Catalina Johnson[1,2], Joshua C. Stein[1], Victor Llaca[3], Kevin Fengler[3], Isabelle Caugant[1], Elodie Belmonte[4], Vincent Pailler[4], Hugo Zalzale[5], Leonardo Pereyra-Bistraín[5], Clément Debiton[4], Pauline Lasserre-Zuber[4], Sophie Bouchet[4], John P. Hamilton[6], Pierre Marin[4], Isabelle Lhommet[4], Hélène Rimbert[4], Michael Alaux[7], Véronique Gautier[4], Pierre Sourdille[4], C. Robin Buell[6], Frédéric Choulet[4], Moussa Benhamed[5], Kellye Eversole[1]

[1]IWGSC, Eau Claire, WI, USA; [2]College of Arts & Sciences, University of Colorado, Boulder, CO, USA; [3]Corteva Agriscience, Johnston, IA, USA; [4]INRAE UMR 1095 GDEC, Clermont-Ferrand, France; [5]Institute of Plant Sciences of Paris-Saclay, Université Paris-Saclay, Orsay, France; [6]Department of Crop & Soil Sciences, Institute of Plant Breeding, Genetics, & Genomics, University of Georgia, Athens, GA, USA; [7]Université Paris-Saclay, INRAE, URGI, Versailles, France; Université Paris-Saclay, INRAE, BioinfOmics, Versailles, France

International initiative to unlock bread wheat diversity by sequencing 12 maximally diverse landrace genomes & update RefSeq Chinese Spring


I W G S C

Wheat Diversity Project

NSF Funded

EU Funded

13 Bread Wheat Accessions

12 Maximally Diverse Landraces
+ 1 cv. Chinese Spring (RefSeq)

Germplasm:
Biological Resources Centre (BRC) for small grains.
Sourdille (INRAE-GDEC)

Genome Sequencing & Assembly

9 Accessions
Corteva Agriscience, Genomics Technologies

4 Accessions
Choulet & Gautier (INRAE-GDEC / Gentyane)

Platinum quality :
- Long-read seq.
- Optical mapping
- Hi-C seq.

Gene Annotation
- Map-based from RefSeq
- De Novo/Evidence-based

Rimbert (INRAE)

Transcriptome Sequencing

All accessions

Benhamed (U. Paris-Saclay)

Gautier (Gentyane/INRAE-GDEC)

mRNA of multiple tissues and developmental stages.

Long-read transcript sequencing

Practical Haplotype Graph
All accessions+
EOT
CoPI Buell (U. Georgia)

IWGSC Data Repository
Alaux (INRAE)

Project leadership & oversight: PI Eversole (IWGSC)
Project management, Education, Outreach & Training (EOT), NCBI liaison: Co-PI Stein

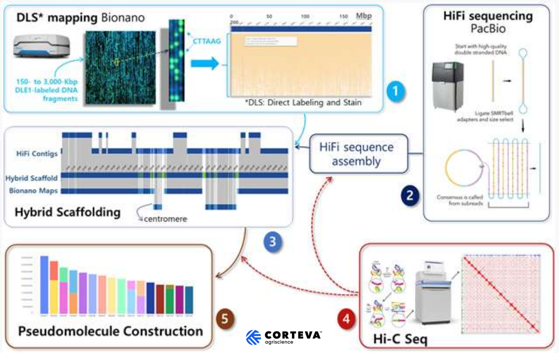
Aim 1) Platinum genome assemblies for 12 maximally diverse landraces and Chinese Spring (CS).

Aim 2) Transcriptome long-reads and gene annotation for 12 landraces while sustaining IWGSC CS RefSeq as a community-curated resource.

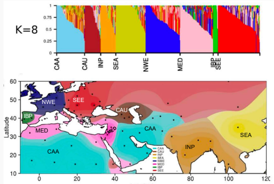
Aim 3) Practical Haplotype Graph (PHG) to catalog all genetic diversity in bread wheat and enable community-driven discoveries across the species;

Aim 4) Increase STEM diversified workforce through education outreach, and training.

Aim 1: Platinum genomes of 12 maximally diverse landraces and cv. Chinese Spring



- 8 phylogeographic haplogroups among 632 genotyped landraces = subpopulations corresponding to early regional growing centers.
- Genome sequence & assembly using long-reads, optical mapping, and conformational capture sequence.



Balfourier et al. 2019

Sequence & Assembly Progress

HiFi read generation:

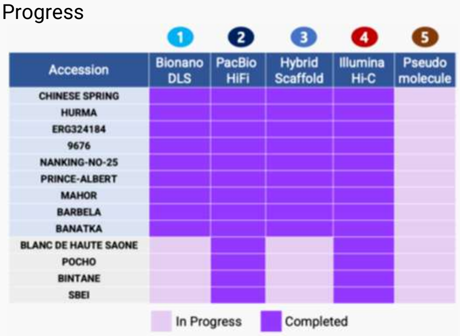
- 6-7 SMRT cells Revio
- 29-40X coverage
- Mean cov. 32.4 ± 2.9X
- Read N50: 17-19 Kb

Hifiasm assembly:

- ~1000 contigs per assembly)
- Contig N50 ~50Mbp
- Max contig length ~340Mbp

DLS map assembly:

- ~600 maps/accession
- Map N50 ~80Mbp
- Map max. len. ~380Mbp
- ~94X molecule coverage

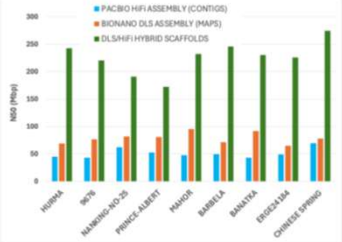


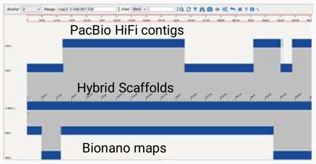
Accession	① Bionano DLS	② PacBio HiFi	③ Hybrid Scaffold	④ Illumina Hi-C	⑤ Pseudo molecule
CHINESE SPRING	■	■	■	■	□
HURMA	■	■	■	■	□
ERG324184	■	■	■	■	□
9676	■	■	■	■	□
NANKING-NO-25	■	■	■	■	□
PRINCE-ALBERT	■	■	■	■	□
MAHOR	■	■	■	■	□
BARBELA	■	■	■	■	□
BANATKA	■	■	■	■	□
BLANC DE HAUTE SAONE	□	■	□	■	□
POCHO	□	■	□	■	□
BINTANE	□	■	□	■	□
SBEI	□	■	□	■	□

□ In Progress ■ Completed

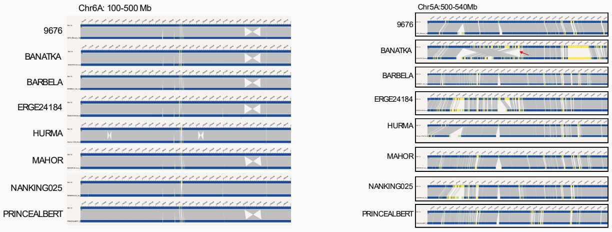
HiFi/DLS hybrid assembly:

- Hybrid scaffold N50 exceeds contig N50 and map N50





Initial comparative genomic views (Bionano maps) show value in revealing conserved and novel structural variation



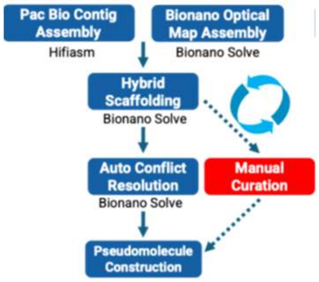
Aim 4: IWGSC undergraduate summer internship in genome curation

3-day training at Corteva Agriscience (Johnston, IA): June 11-13, 2024

- Principles of optical mapping, long read sequencing, assembly, and chromosome reconstruction.
- Training on Bionano Access software, visualization and interpretation of assembly data.
- Informative tours of Corteva's core labs and greenhouse facilities.

Genome Curation Project

- Automated conflict resolution is good but imperfect
- Manual curation often improves assembly accuracy
- But manual curation is time consuming and costly


Pac Bio Contig Assembly
Hifiasm

Bionano Optical Map Assembly
Bionano Solve

Hybrid Scaffolding
Bionano Solve

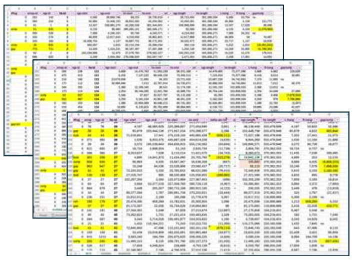
Auto Conflict Resolution
Bionano Solve

Manual Curation

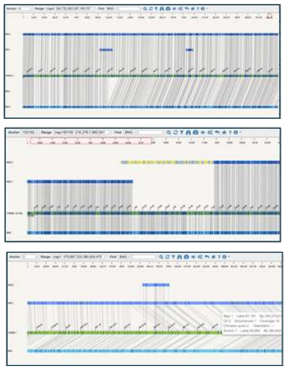
Pseudomolecule Construction

Pre-screen:

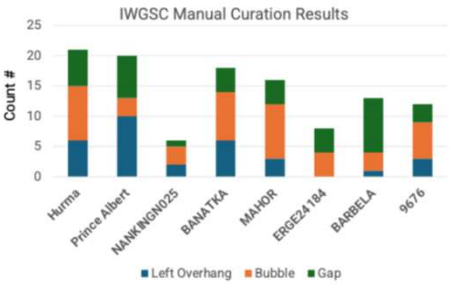
- Perl script reads XMAP output files
- FLAGS potential problems
- Gaps, overlaps, and overhangs
- Output loaded into spreadsheets



Visual Inspection:



- Bionano hybrid scaffold browser
- Checks flagged regions
- Notates as OK or needing further manual inspection/intervention



Project Funding

NSF Award #2322957
ERC Project 3Dwheat #101044399
INRAE Plant Biology and Breeding Department
IWGSC

IWGSC Sponsors

References

Balfourier F, et al. International Wheat Genome Sequencing Consortium; BreedWheat Consortium; Paux, E. Worldwide Phylogeography and History of Wheat Genetic Diversity. *Sci Adv* **2019**, *5* (5), eaav0536.

Zhu T, et al. Optical Maps Refine the Bread Wheat Triticum Aestivum Cv. Chinese Spring Genome Assembly. *Plant J.* **2021**, *107* (1), 303–314.