

IWGSC 2.0: Wheat Genome Curation for Future Crop Resilience

The International Wheat Genome Sequencing Consortium¹, Catalina Johnson^{1,2}, Joshua C. Stein¹, Victor Llaca³, Kevin Fengler³, Isabelle Caugant¹, Elodie Belmonte⁴, Vincent Pailler⁴, Hugo Zalzale⁵, Leonardo Pereyra-Bistraín⁵, Clément Debiton⁴, Pauline Lasserre-Zuber⁴, Sophie Bouchet⁴, John P. Hamilton⁶, Pierre Marin⁴, Isabelle Lhommet⁴, Hélène Rimbart⁴, Michael Alaux⁷, Véronique Gautier⁴, Pierre Sourdille⁴, C. Robin Buell⁶, Frédéric Choulet⁴, Moussa Benhamed⁵, Kellye Eversole¹

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

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

Wheat Diversity Project

13 Bread Wheat Accessions
12 Maximally Diverse Landraces + 1 cv. Chinese Spring (RefSeq)

Genome Sequencing & Assembly

- 9 Accessions: Corteva Agriscience, Genomics Technologies
- 4 Accessions: Choulet & Gautier (INRAE-GDEC / Gentyane)

Gene Annotation

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

Practical Haplotype Graph

All accessions+ EOT

IWGSC Data Repository

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

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

DLS+ mapping Bionano

HiFi sequencing PacBio

HiFi sequence assembly

Hybrid Scaffolding

Pseudomolecule Construction

Hi-C Seq

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.

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	Bionano DLS	PacBio HiFi	Hybrid Scaffold	Illumina Hi-C	Pseudo molecule
CHINESE SPRING	In Progress	Completed	Completed	Completed	Completed
HURMA	In Progress	Completed	Completed	Completed	Completed
ERG24184	In Progress	Completed	Completed	Completed	Completed
9676	In Progress	Completed	Completed	Completed	Completed
NANKING-NO-25	In Progress	Completed	Completed	Completed	Completed
PRINCE-ALBERT	In Progress	Completed	Completed	Completed	Completed
MAHOR	In Progress	Completed	Completed	Completed	Completed
BARBELA	In Progress	Completed	Completed	Completed	Completed
BANATKA	In Progress	Completed	Completed	Completed	Completed
BLANC DE HAUTE SAONE	In Progress	Completed	Completed	Completed	Completed
POCHO	In Progress	Completed	Completed	Completed	Completed
BINTANE	In Progress	Completed	Completed	Completed	Completed
SBEI	In Progress	Completed	Completed	Completed	Completed

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.

HiFi/DLS hybrid assembly:

- Hybrid scaffold N50 exceeds contig N50 and map N50

Genome Curation Project

Pre-screen:

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

Workflow: Pac Bio Contig Assembly (Hifiasm) and Bionano Optical Map Assembly (Bionano Solve) feed into Hybrid Scaffolding (Bionano Solve). This leads to Auto Conflict Resolution (Bionano Solve) and Manual Curation. The final step is Pseudomolecule Construction.

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

Visual Inspection:

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

IWGSC Manual Curation Results

Project Funding

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

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.