I W G S C
International Wheat Genome
Sequencing Consortium

IWGSC 2.0 Implementing the Wheat Diversity Project

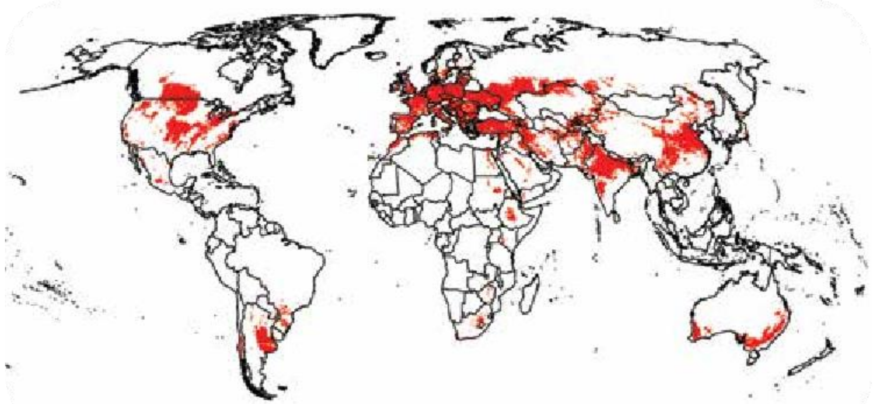
Catalina Johnson
IWGSC

PAG 32

Plant and Animal Genome Conference
San Diego, CA, USA
14 January 2025

Bread Wheat is a Global Food Security Crop

- 750 million metric tons per year global production
- Provides ~20% human dietary calories and protein
- More nutritionally complete than other cereal grains: Protein, B vitamins, and iron.



Lantican, M.A., H.J. Braun, T.S. Payne, R.P. Singh, K. Sonder, M. Baum, M. van Ginkel, and O. Erenstein. 2016. Impacts of International Wheat Improvement Research, 1994-2014. Mexico, D.F.: CIMMYT.

https://upload.wikimedia.org/wikipedia/commons/c/c7/Korb_mit_Brötchen.JPG [3268zauber]

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

IWGSC History & Scientific Impacts

2005 vision: high quality bread wheat genome sequence to empower all aspects of basic and applied wheat science.

Large hexaploid genome (16-Gb)
5x Human genome; 42x Rice genome
~108,000 protein-coding genes

Timeline to the first bread wheat IWGSC RefSeq - Chinese Spring

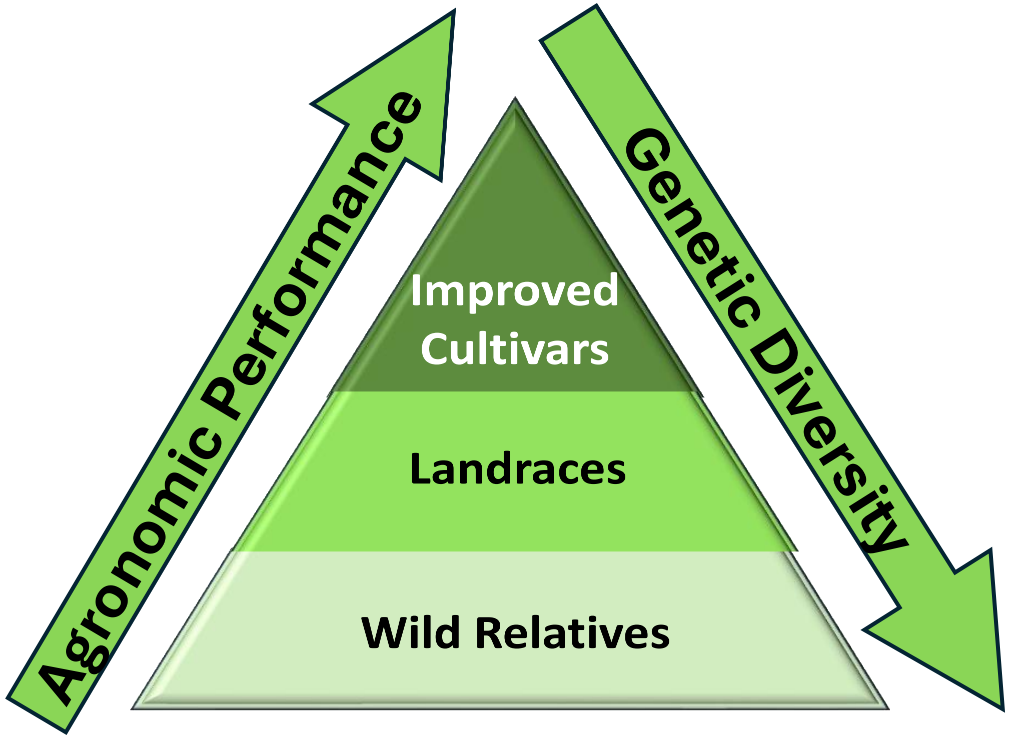
2021: Publication/release of IWGSC RefSeq v2.1 and annotation
2018: Publication and analysis of IWGSC RefSeq v1.0
2017: Pre-publication release of IWGSC RefSeq v1.0
2016: Chromosome-based whole genome shotgun assembly (IWGSC WGA v0.4).
2015: Physical maps for all chromosomes completed
2014: Publication of first chromosome based draft genome sequence
2005: IWGSC formed by wheat growers, plant scientists, and public and private breeders

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Breeder’s Dilemma

- **Year-over-year increase production is required.**
- **Climate change promising to destabilize production trends.**
- **Erosion of genetic diversity.**
- **Genetic gains are sluggish; TFP is .7% vs. annual target of 2.03%**

(Dixon J. Wheat Facts and Futures: CIMMYT; 2009. p. 1-19; Rosegrant Global Food Demand, Supply, and Price Prospects to 2010: International Food Policy Research Institute; 2010.)



Agronomic Performance

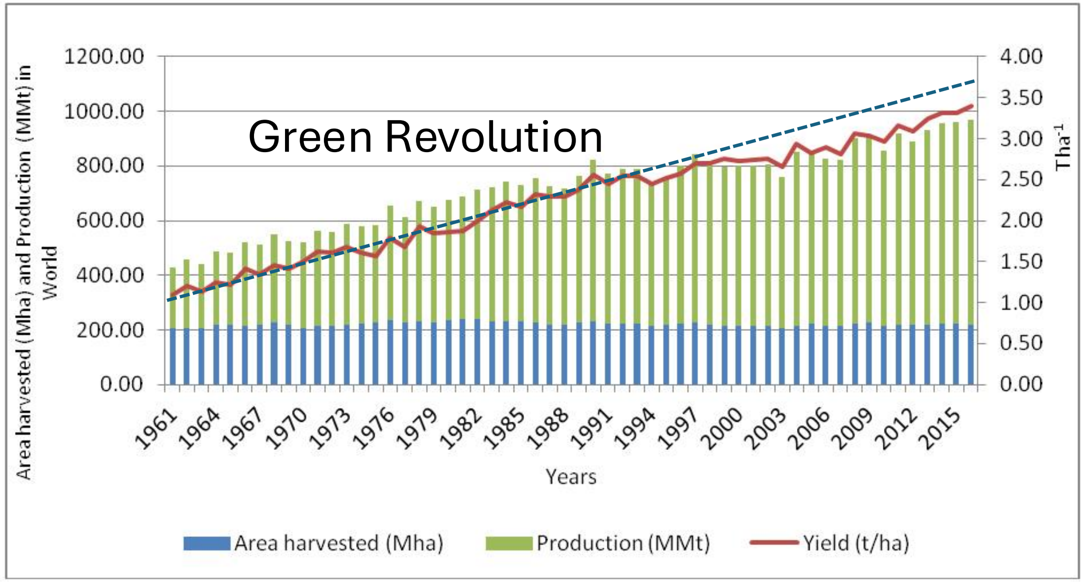
Genetic Diversity

Improved Cultivars

Landraces

Wild Relatives

World wheat yield, production and area from 1961 to 2016 (FAO, 2018)



Tadesse et al. 2019. Crop Breed Genet Genom. 2019;1:e190005.

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Landraces as a Foundational Source of Wheat Diversity

- Locally-adapted varieties (landraces) farmed for thousands of years before displacement by modern varieties – at risk of extinction, preserved at INRAE.

- Genetic diversity is mostly absent in modern varieties, making these landraces valuable trait reservoirs.


~7000 YBP

Domestication ~10,000 YBP

~5000 YBP


ERGE7966 (Northwest Europe)

© INRA, Lionel BARDY

INRAE Small Grains BRC

Resistance to leaf rust (*Puccinia triticina*)

Zhao, X., Guo, Y., Kang, L. *et al*. Population genomics unravels the Holocene history of bread wheat and its relatives. *Nat. Plants* **9**, 403–419 (2023). https://doi.org/10.1038/s41477-023-01367-3
Wang, Z.;et al. . Dispersed Emergence and Protracted Domestication of Polyploid Wheat Uncovered by Mosaic Ancestral Haploblock Inference. *Nat. Commun.* **2022**, *13* (1), 3891.

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

IWGSC Wheat Diversity Project

Preserve, discover, and mobilize genetic resources by sequencing landrace genomes that encompass the worldwide diversity of bread wheat.

- **Aim 1: Generate platinum quality genome assemblies for 12 maximally diverse bread wheat landrace accessions and the community reference cv. Chinese Spring (CS).**
- **Aim 2: Establish transcriptome data and gene annotations for these accessions and update the IWGSC CS RefSeq annotation in a manner that sustains its role as a community- curated resource.**
- **Aim 3: Construct a Practical Haplotype Graph (PHG) to catalog all genetic diversity within bread wheat and enable community-driven discoveries across the species.**

Principal Investigator

Kellye Eversole

Biological Resources, Genome Production, Annotation

Pierre Sourdille

Frédéric Choulet

Hélène Rimbert

Véronique Gautier

IWGSC Data Repository

Michael Alaux

Transcriptome

Moussa Benhamed

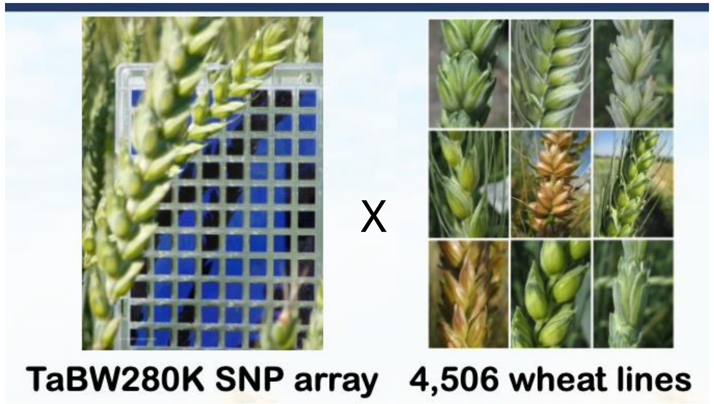
Practical Haplotype Graph

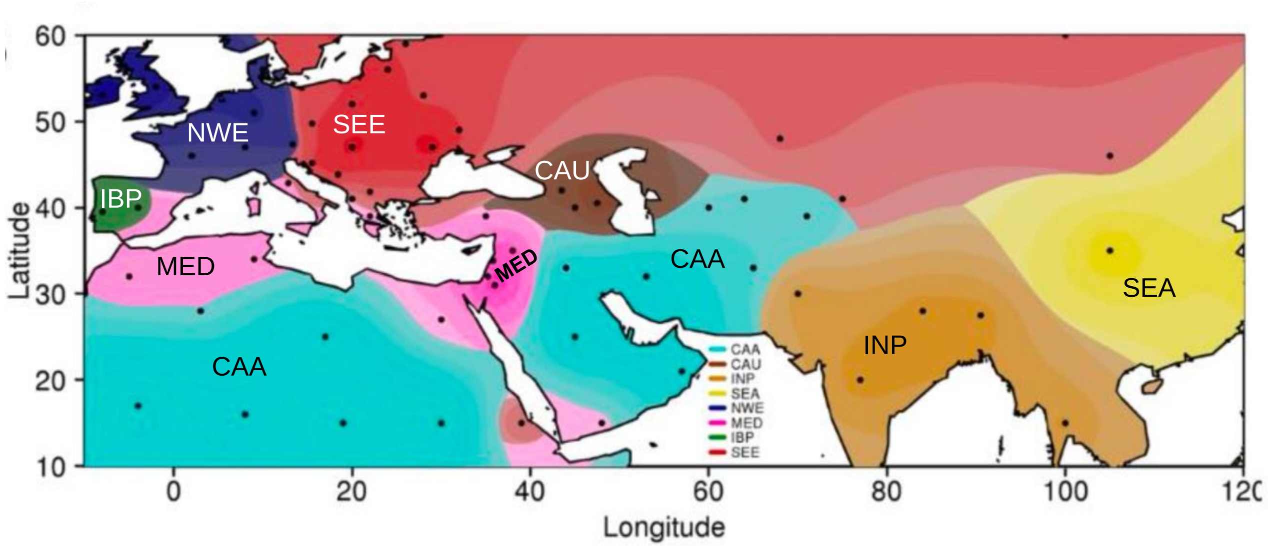
C. Robin Buell

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Selection of 8 Foundational Phylogeographic Haplotypes

- Sampled 4506 wheat lines from 105 countries, including 632 Landraces, preserved at the INRAE Small Grains Biological Research Centre.
- Identified 8 Phylogeographic Haplogroups among landraces.
- Selected 12 maximally diverse lines showing minimal admixture.



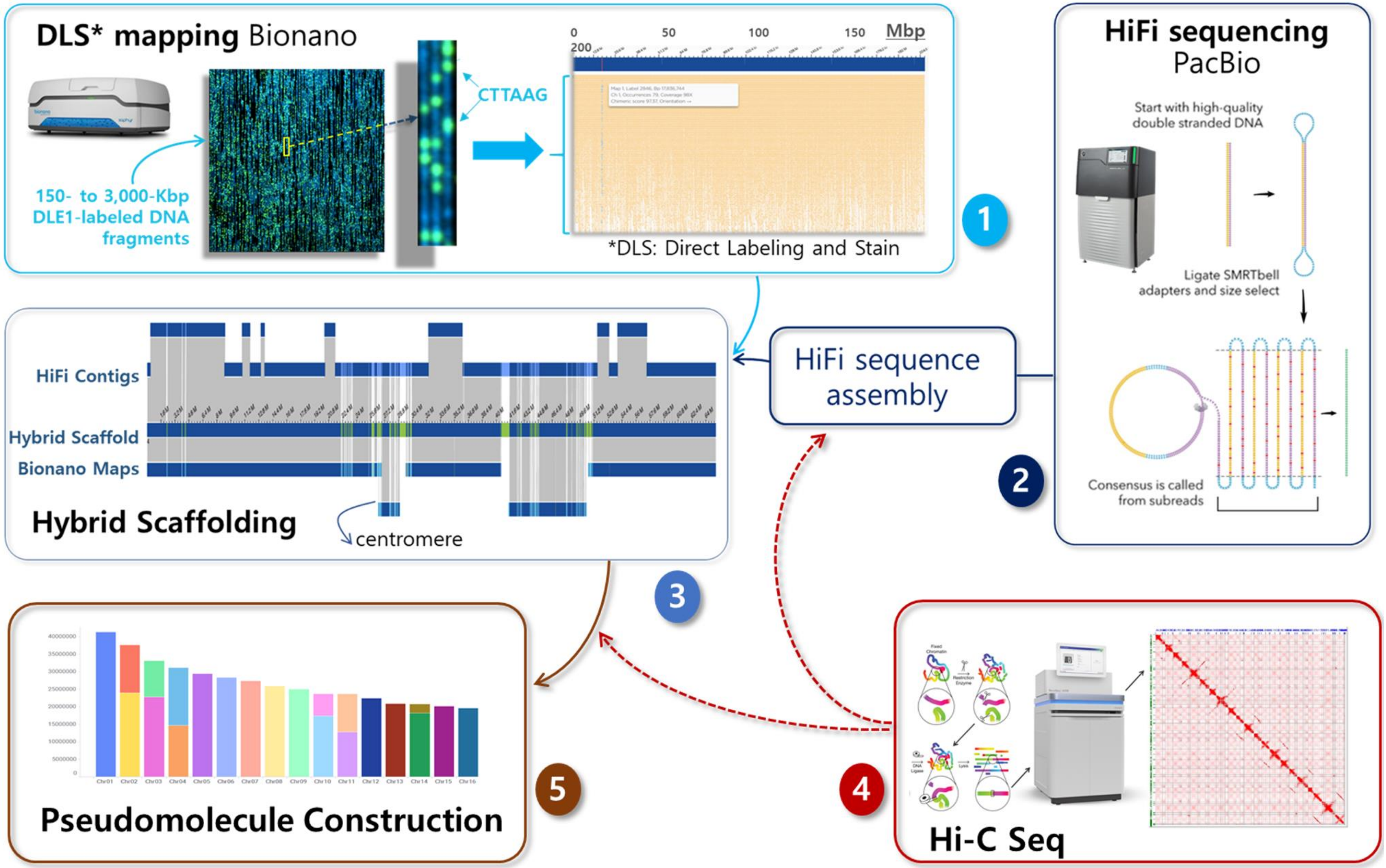


	Haplogroup	Country	Habit	Assign-8†	Assign-12†
	CAA	Iraq	Spring	0.994	0.892
	CAU	Palestine	Winter	1	1
	INP	India	Spring	1	1
	SEA	China	Winter	1	1
	NWE	France	Winter	1	1
	MED	Tunisia	Spring	1	1
	IBP	Portugal	Spring	1	1
	SEE	Ukraine	Winter	1	1
	NWE	Turkey	Spring	0.676	1
E*	NWE	France	Winter	0.506	1
	SEE	Morocco	Spring	0.714	1
	NA	Colombia	Spring	NA	NA

Balfourier, F., Bouchet, S., Robert, S., De Oliveira, R., Rimbert, H., Kitt, J., Choulet, F., International Wheat Genome Sequencing Consortium, BreedWheat Consortium, & Paux, E. (2019). Worldwide phylogeography and history of wheat genetic diversity. *Science Advances*, *5*(5), eaav0536. https://doi.org/10.1126/sciadv.aav0536

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

IWGSC Wheat Diversity Project Strategy



DLS* mapping Bionano

150- to 3,000-Kbp DLE1-labeled DNA fragments

*DLS: Direct Labeling and Stain

HiFi sequencing PacBio

Start with high-quality double stranded DNA

Ligate SMRTbell adapters and size select

Consensus is called from subreads

HiFi sequence assembly

Hybrid Scaffolding

Hi-C Seq

Pseudomolecule Construction

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

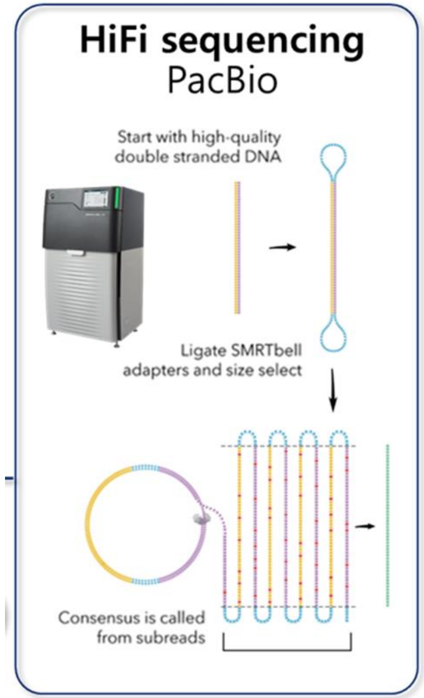
Progress on Data Generation

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

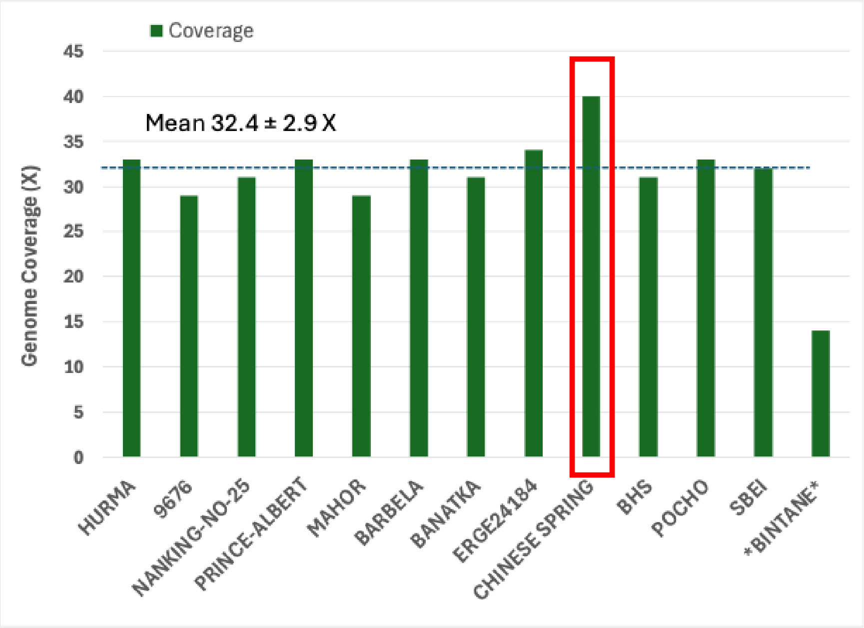
A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

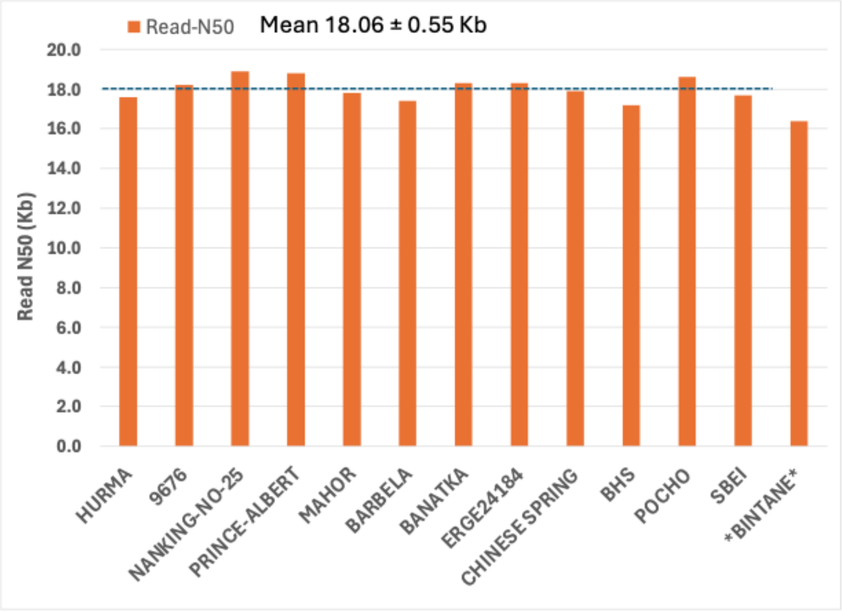
PacBio HiFi Progress



- Target: 25-35X coverage (6-7 SMRTcells Revio)
- Status: completed 12 accessions; 1 in-progress
- Achieved: 29-40X coverage (mean 32.4 ± 2.9X)
- Read N50: 17-19 Kb

Special attention to Chinese Spring: 40X coverage

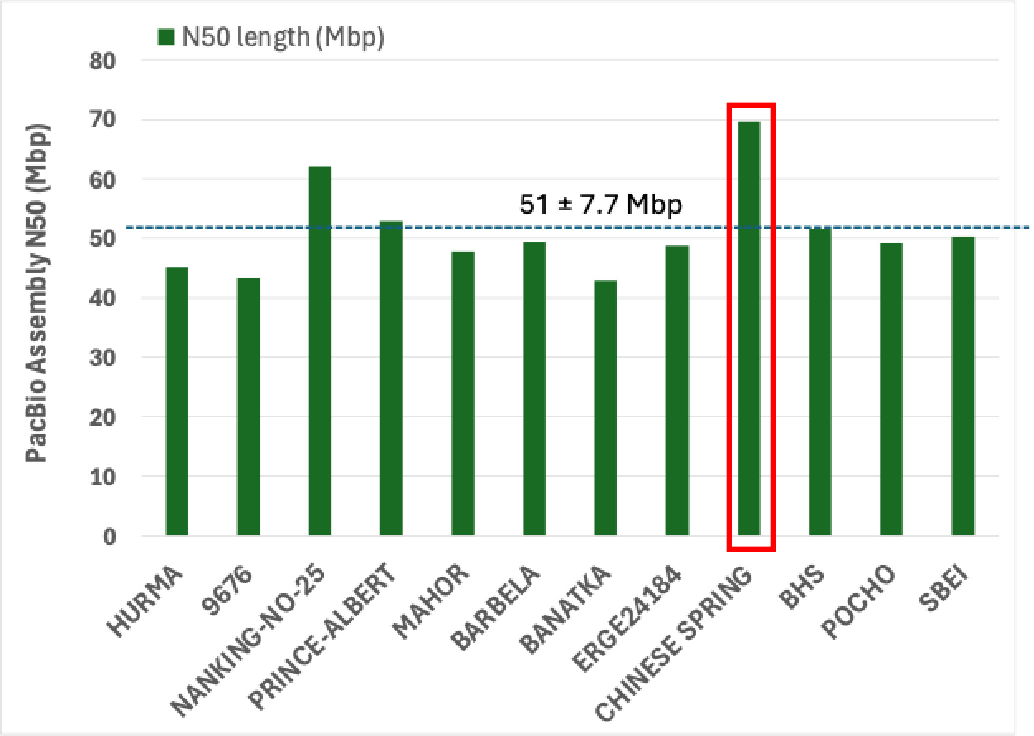




A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

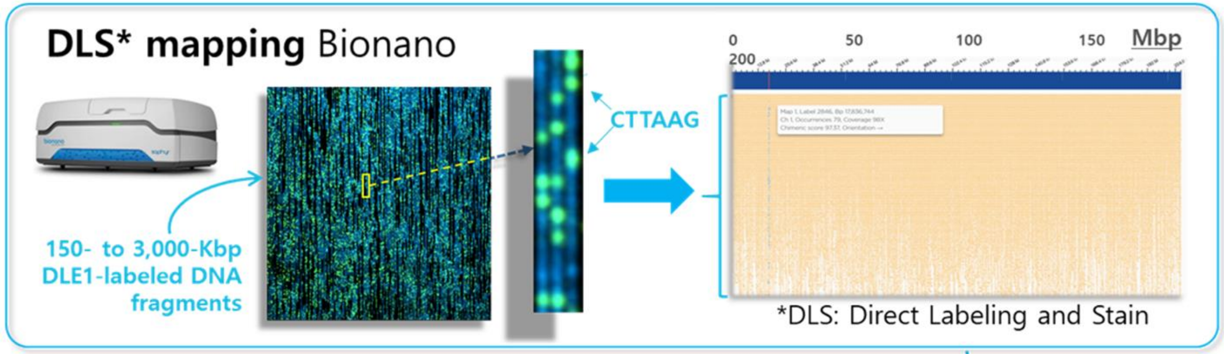
PacBio HiFi Assembly

- Contig Assembly: Hifiasm
- ~1000 contigs
- N50 ~ 50Mbp
- Max ~340Mbp



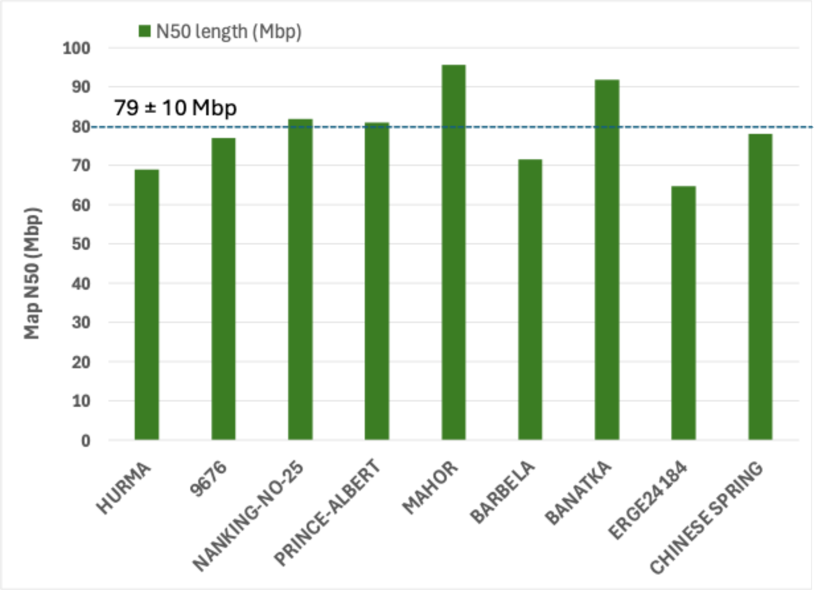
A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

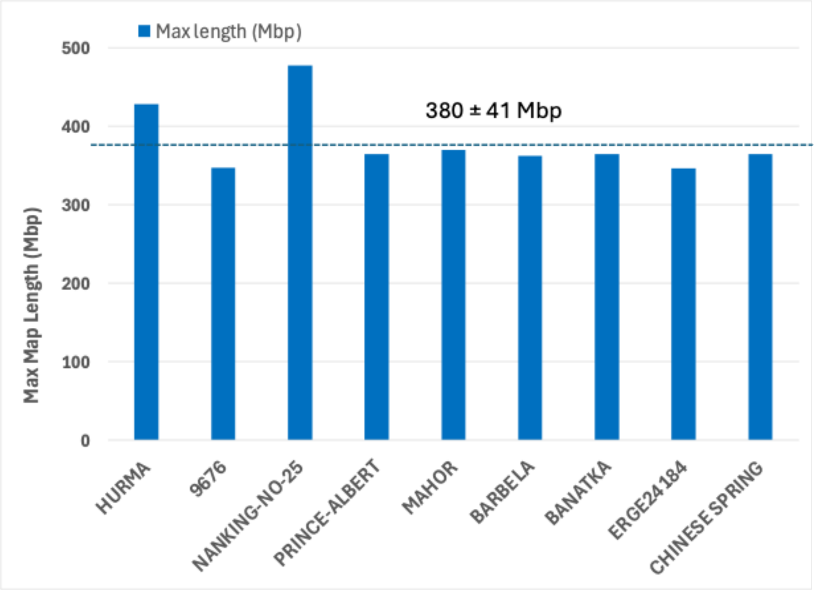
Bionano Map Generation

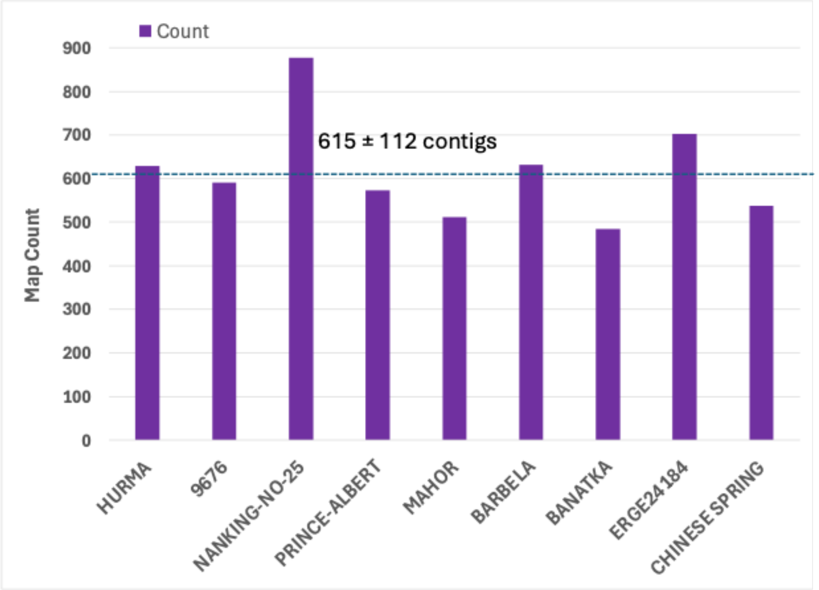


*DLS: Direct Labeling and Stain

- ~600 maps/accession
- N50 ~80Mbp
- Max ~380 Mbp
- ~94X molecule coverage

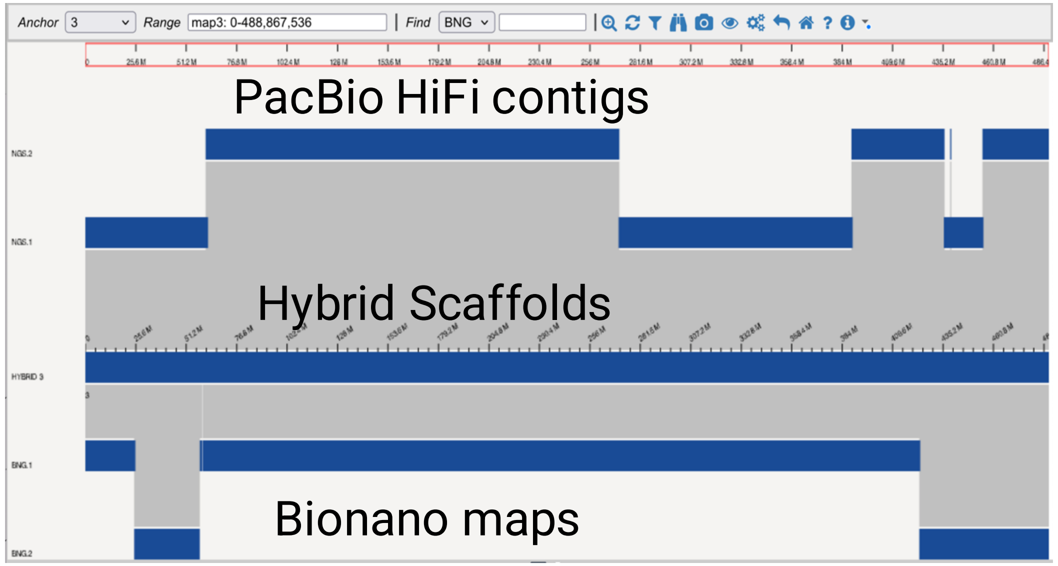






A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

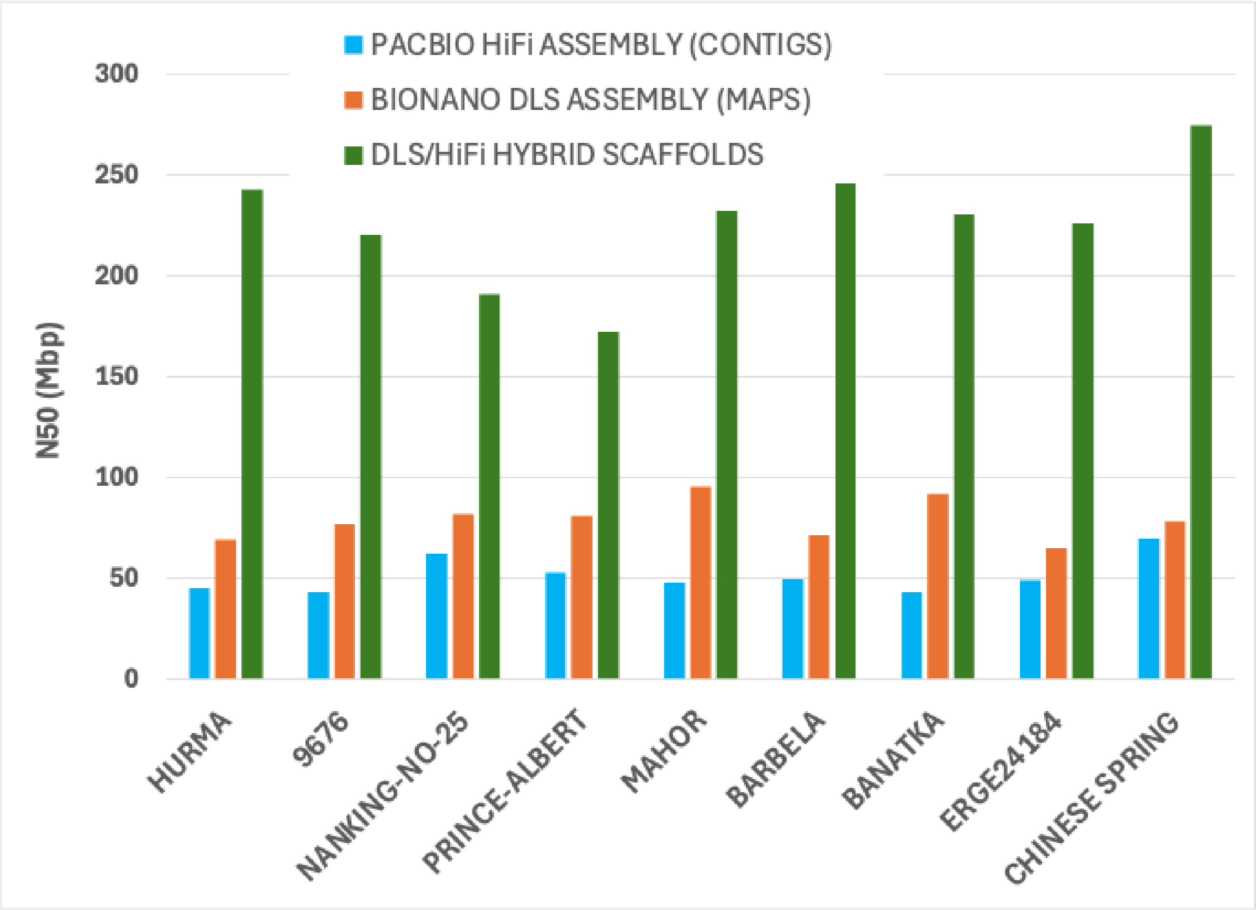
Hybrid Scaffolds



Anchor 3

Range map3: 0-488,867,536

Find BNG



Advantages:

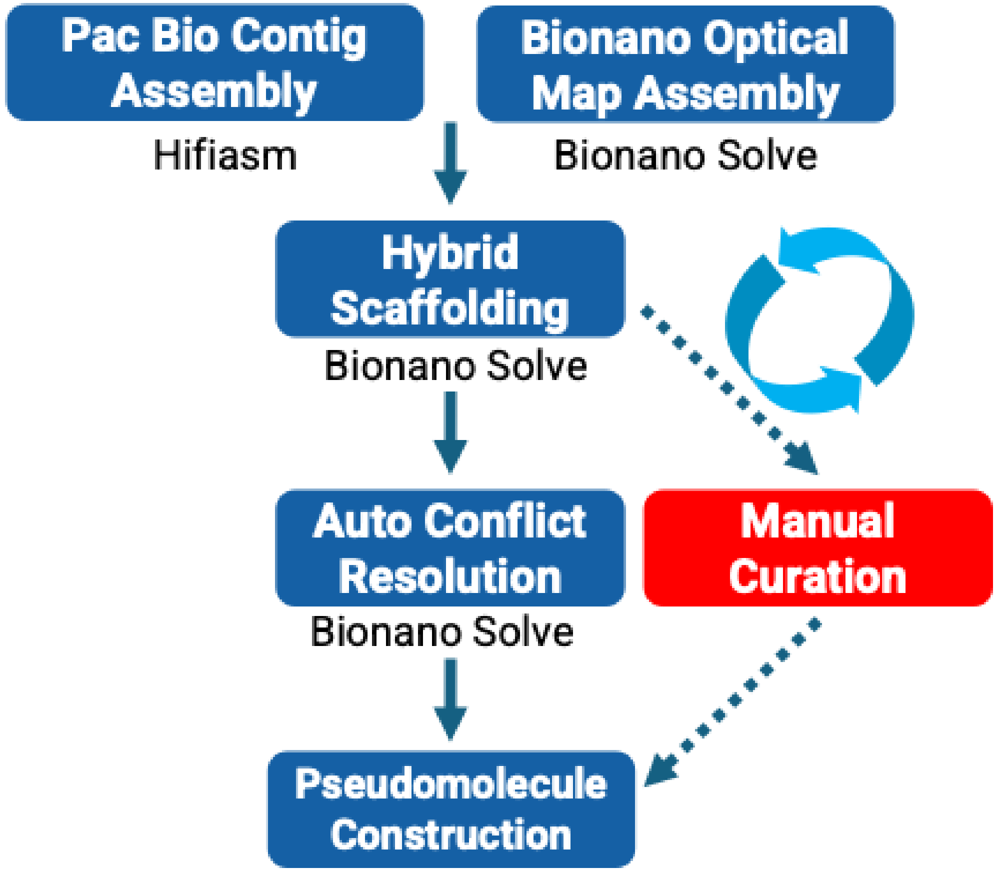
1) Orders and orients contigs in physical space
2) Accurate gap sizing
3) Genome and assembly characterization
4) Auto conflict resolution of chimeric contigs
5) Auto overlapping contig “resolution”

Visual inspection is key to manual curation, identify issues with contigs or maps

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Manual Curation

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

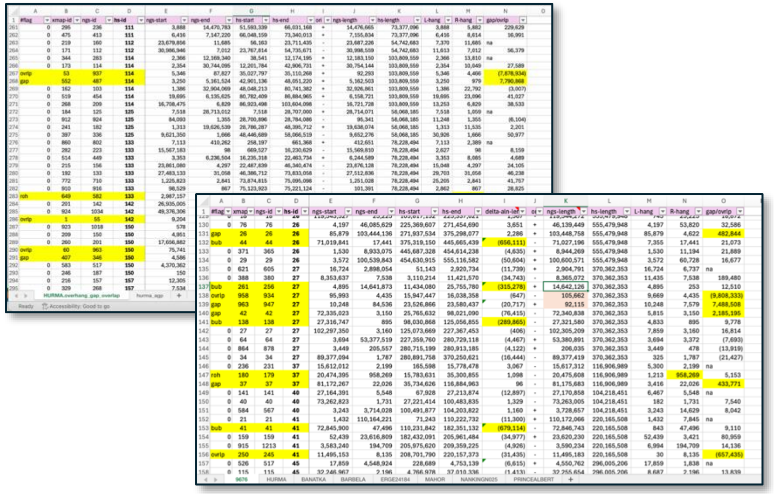


A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Aim 4: IWGSC undergraduate summer internship in genome curation

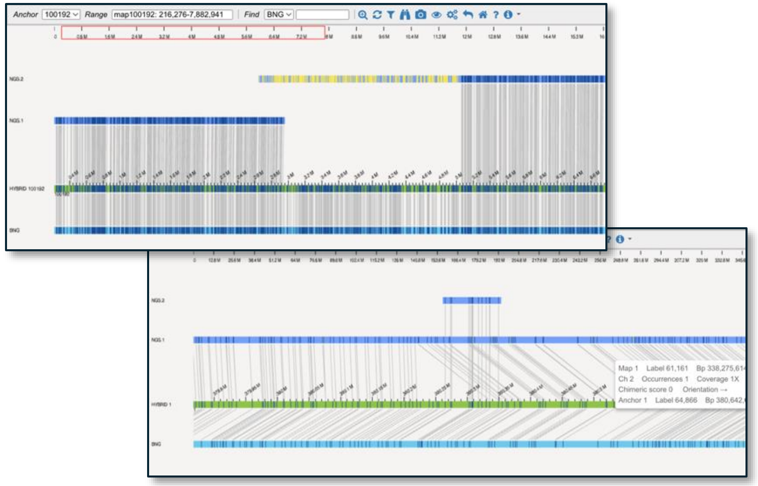
Pre-screen:

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



Visual Inspection:

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

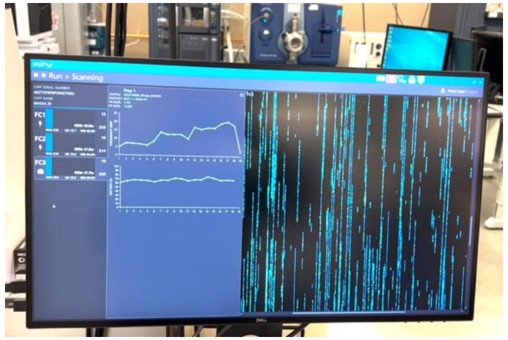


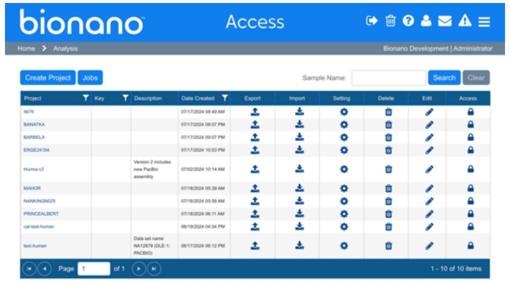
A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

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.





Project	Key	Description	Date Created	Export	Import	Setting	Delete	Edit	Access
9676			07/17/2024 08:49 AM	⬆	⬇	⚙	🗑	✎	🔒
BANATKA			07/17/2024 08:07 PM	⬆	⬇	⚙	🗑	✎	🔒
BARBELA			07/17/2024 09:07 PM	⬆	⬇	⚙	🗑	✎	🔒
ERGE24184			07/17/2024 10:53 PM	⬆	⬇	⚙	🗑	✎	🔒
Hurma-v2		Version 2 includes new PacBio assembly	07/02/2024 10:14 AM	⬆	⬇	⚙	🗑	✎	🔒
MAHOR			07/18/2024 05:39 AM	⬆	⬇	⚙	🗑	✎	🔒
NANKINGN025			07/18/2024 05:56 AM	⬆	⬇	⚙	🗑	✎	🔒
PRINCEALBERT			07/18/2024 06:11 AM	⬆	⬇	⚙	🗑	✎	🔒
cat-test-human			06/19/2024 04:34 PM	⬆	⬇	⚙	🗑	✎	🔒
test-human		Data set name NA12878 (DLE-1; PACBIO)	06/17/2024 06:12 PM	⬆	⬇	⚙	🗑	✎	🔒

Page 1 of 1

1 - 10 of 10 items

Victor Llaca

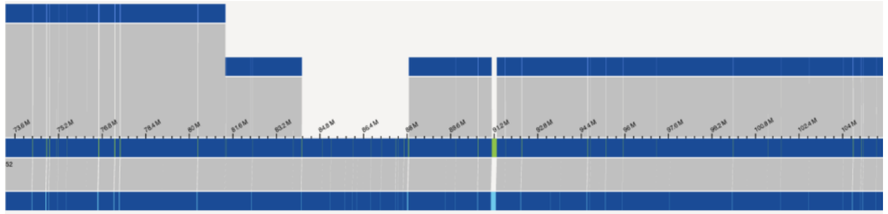
Kevin Fengler

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

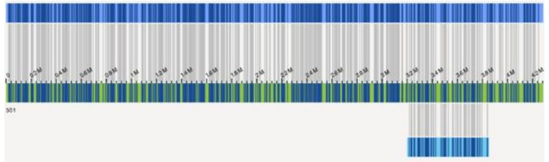
Features that need manual curation

Gaps in the hybrid scaffold

Super-Scaffold_52

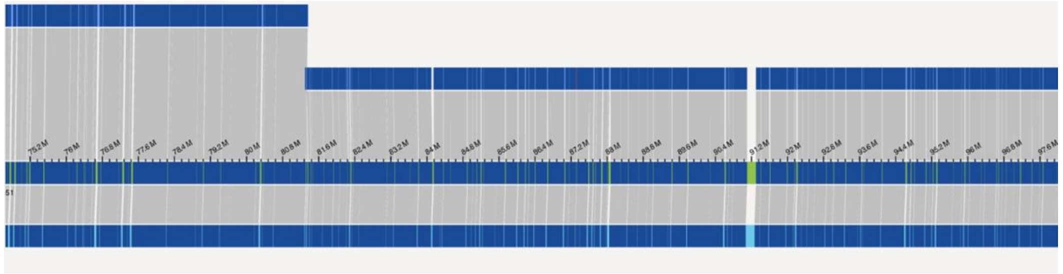


Super-Scaffold_501 (4.2 Mb scaffold)



Low coverage/het map

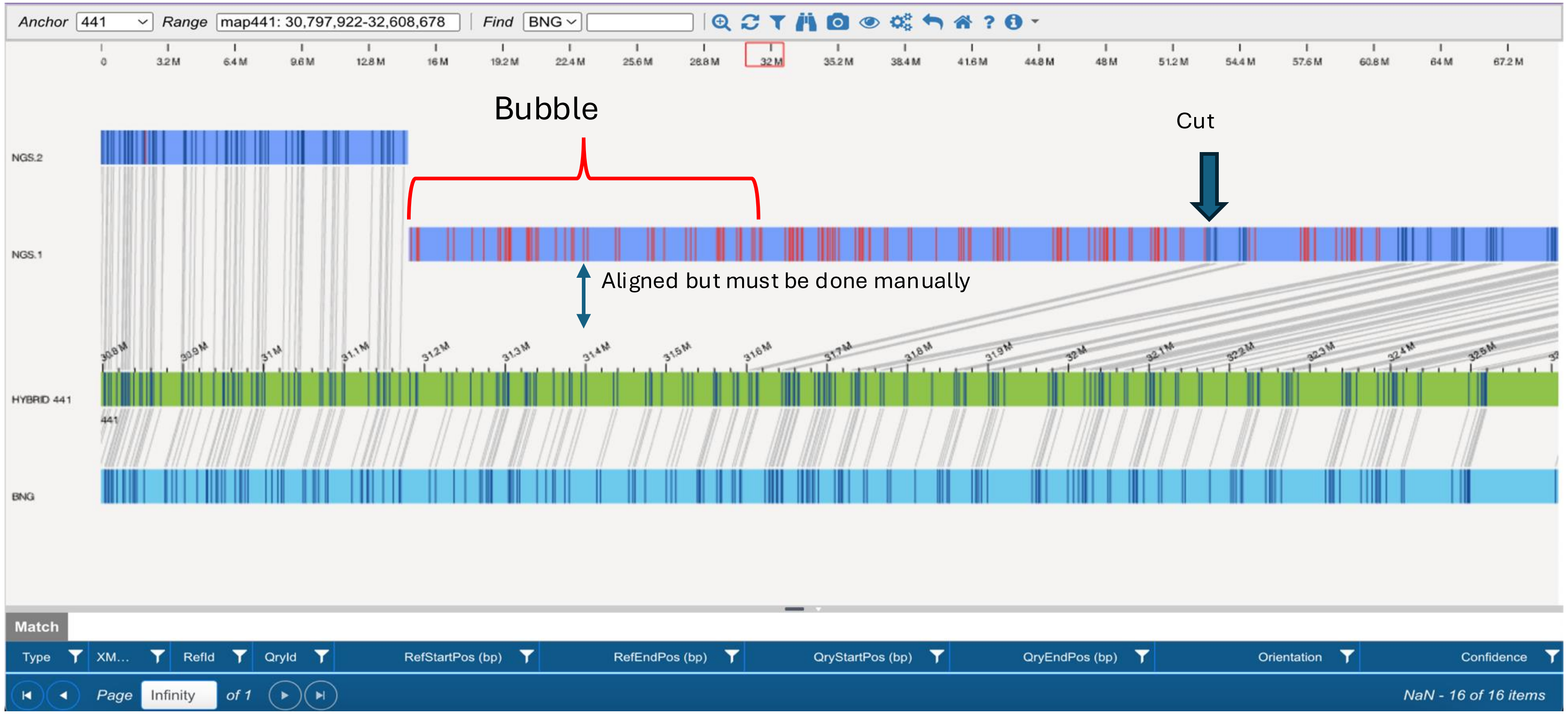
Remove specific problematic map
Or
Filter for small/low-coverage maps



A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Features that need manual curation

“Bubbles”



Anchor 441 Range map441: 30,797,922-32,608,678 Find BNG

Bubble

Cut

Aligned but must be done manually

Type	XM...	RefId	QryId	RefStartPos (bp)	RefEndPos (bp)	QryStartPos (bp)	QryEndPos (bp)	Orientation	Confidence

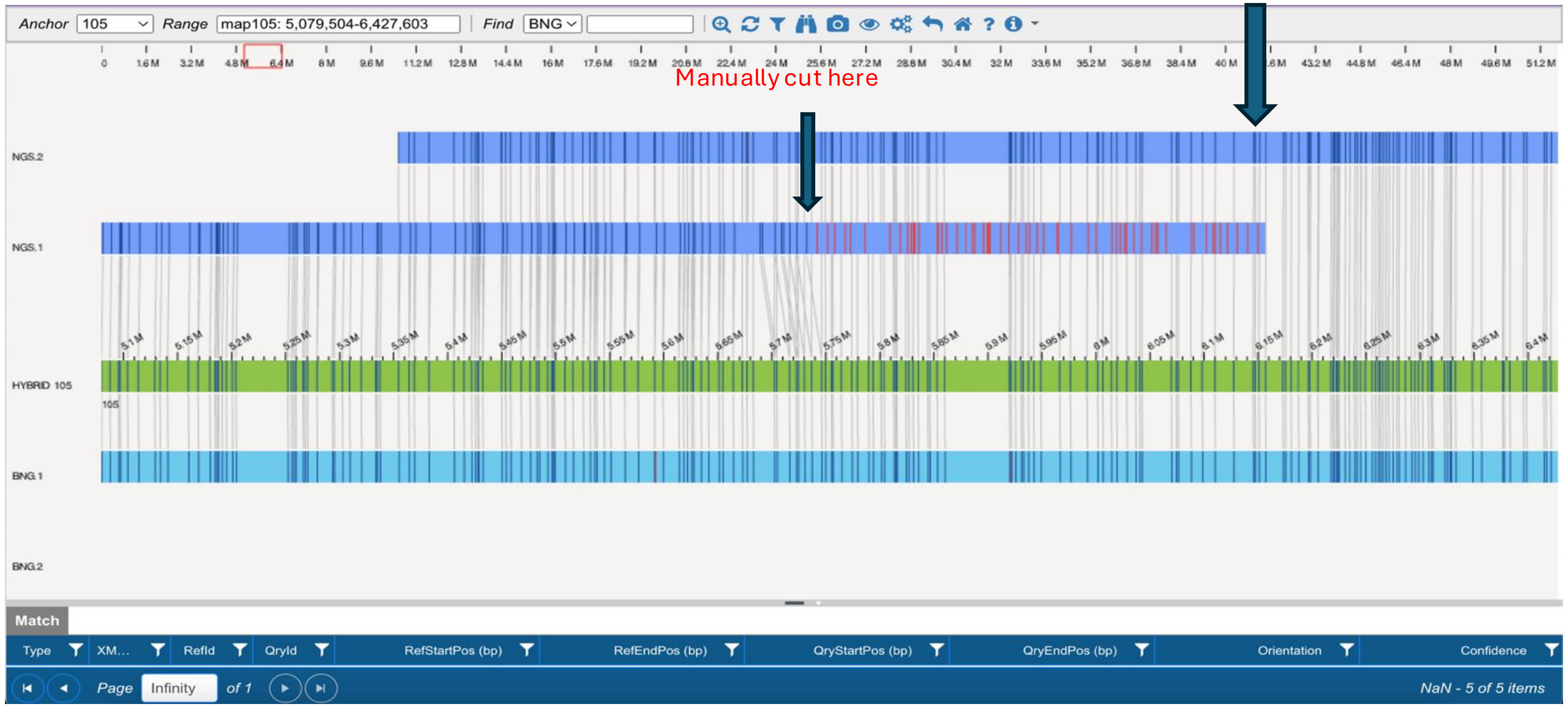
Page Infinity of 1

NaN - 16 of 16 items

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Features that need manual curation

Overlap not trimmed as wanted


Instead of automatic cut here

Manually cut here

Anchor 105 Range map105: 5,079,504-6,427,603 Find BNG

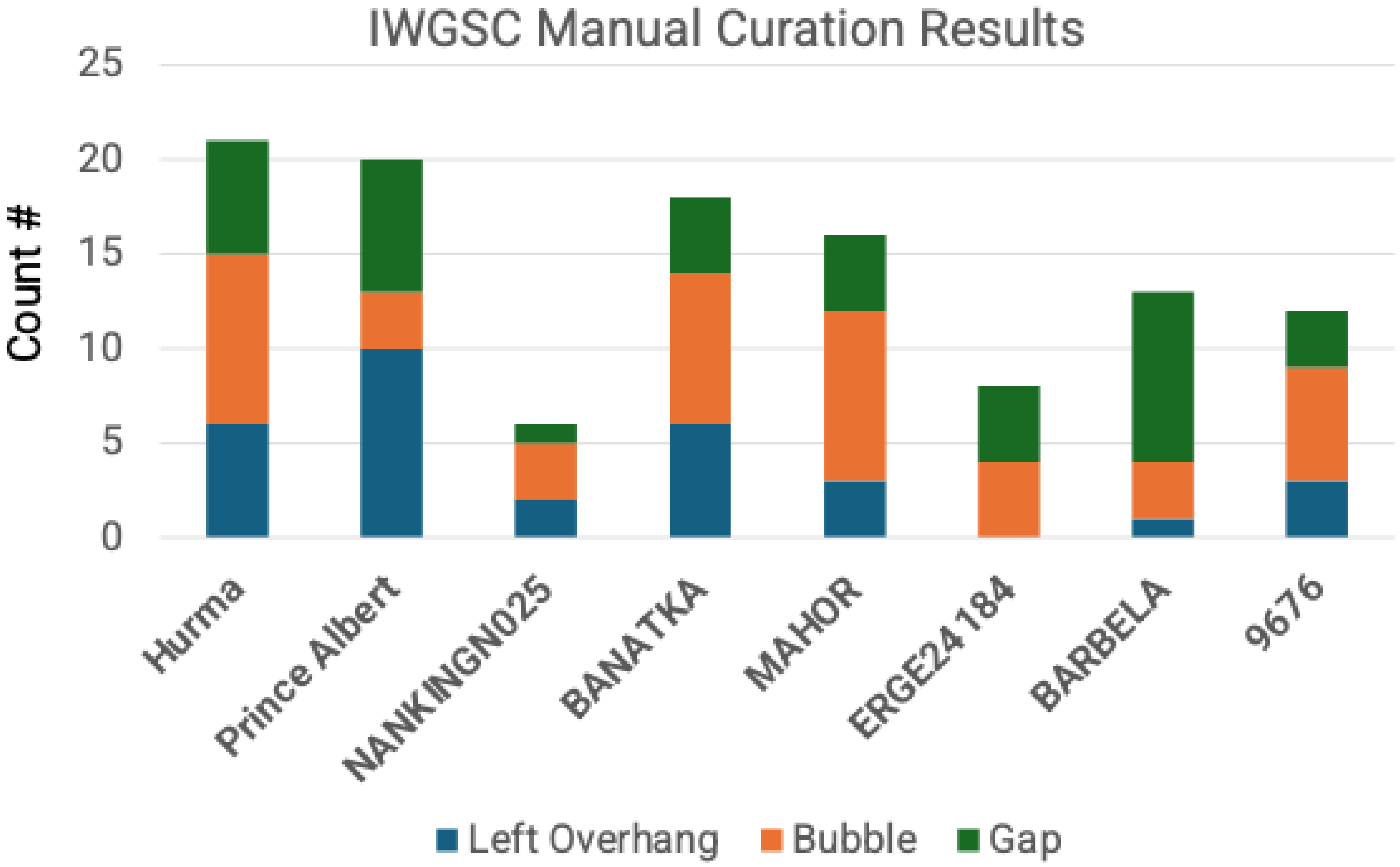
Type	XM...	RefId	QryId	RefStartPos (bp)	RefEndPos (bp)	QryStartPos (bp)	QryEndPos (bp)	Orientation	Confidence

Page Infinity of 1

NaN - 5 of 5 items

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Features that need manual curation



A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Conclusions

- Bread wheat landraces are a rich sources of genetic diversity largely absent in modern breeding materials & reservoir for adaptive traits.
- PHG of 12 platinum-quality genome sequences sampled from 8 foundational phylogenomic haplogroups will generate robust genotyping platform to dissect agronomic traits and accelerate breeding.
- Stewardship of the updated IWGSC RefSeq + Annotation will sustain this scientific resource into the future.
- All landrace germplasm publicly available through INRAE small grains BRC.
- Through my undergraduate internship opportunity, I have collaborated with top scientists of the field and learned cutting edge technologies.

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Contact information:
catjo99@gmail.com | Follow me on linkedin:

About me:

- Senior at University of Colorado Boulder
- Majoring in Ecology and Evolutionary Biology
- Minoring in Computational Biology
- Working in the Quandt Mycology Lab and the Suding Grassland Restoration Lab
- Honors Thesis: Plant-soil feedbacks and grassland competition in the Front Range of Colorado

Plans:

- Work in industry before pursuing a PhD!
- Interests are microbial ecology, mycology, computational biology, and agriculture

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Wheat Diversity Contributors

INRAE-GDEC/Gentyane
Pierre Sourdille
Frédéric Choulet
Hélène Rimbert
Véronique Gautier
Elodie Belmonte
Vincent Pailler
Clément Debiton
Pauline Lasserre-Zuber
Pierre Marin
Sophie Bouchet

Corteva Agriscience
Victor Llaca
Kevin Fengler

IWGSC
Kellye Eversole
Joshua Stein
Catalina Johnson

INRAE-URGI
Michael Alaux

U. Paris-Saclay/IPS2
Moussa Benhamed
Hugo Zalzale
Leonardo Pereyra-Bistraín

U. Georgia, Athens
Robin Buell
John P. Hamilton

A project supported by the National Science Foundation under grant #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the International Wheat Genome Sequencing Consortium

Funding Acknowledgments

National Science Foundation grant #2322957

Funded by the European Union

European Research Council
Established by the European Commission

European Research Council ERC (project 3Dwheat #101044399)

INRAE Plant Biology and Breeding department

International Wheat Genome Sequencing Consortium