The IWGSC Wheat Diversity Project: Foundational Genome Resources to Advance a Global Food Security Crop

The International Wheat Genome Sequencing Consortium[1], Joshua Stein[1], Gina Zastrow-Hayes[2], Victor Llaca[2], Hélène Rimbert[3], Frédéric Choulet[3], Clément Debiton[3], Charles Poncet[3], Michael Alaux[4], Pierre Sourdille[3], Moussa Benhamed[5], C. Robin Buell[6], and Kellye Eversole[1]

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

This project addresses the critical need to preserve, discover, and mobilize genetic resources by sequencing landrace genomes that encompass the worldwide diversity of bread wheat.

ABSTRACT:

Crop diversity is essential to drive the future adaptive capacity of agriculture. In bread wheat (*Triticum aestivum* L. $2n = 6x = 42$, AABBDD), a significant amount of diversity is held within landraces traditionally farmed across Northern Africa, Europe, and Asia for thousands of years, and, while the varieties are no longer grown today, they remain a key reservoir for adaptive traits. Today's high-yielding varieties contain only a small fraction of this available gene pool, however, raising concerns that genetic erosion within modern stocks could limit future gains in wheat production. Further, this resource could contain traits that would help modern varieties meet the challenges of production in changing climates and the accompanying disease and pest pressures. This International Wheat Genome Sequencing Consortium (IWGSC) Wheat Diversity Project addresses the critical need to preserve, discover, and mobilize these genetic resources by sequencing 12 landrace genomes that are foundational to early regional growing centers and represent the worldwide diversity of bread wheat. Platinum-quality reference genomes produced by this project will also update the IWGSC Chinese Spring reference sequence (IWGSC CS RefSeq), which serves as a primary reference sequence for bread wheat research. Comprehensive annotation and representation of the wheat pangenome as a Practical Haplotype Graph will provide powerful tools to accelerate future genomics and breeding research. This project is funded by NSF Award #2322957, the European Research Council ERC (project 3Dwheat #101044399), INRAE Plant Biology and Breeding department, and the IWGSC.

International Wheat Genome Sequencing Consortium (IWGSC) Wheat Diversity Project

Key: NSF Funded / EU Funded

Oversight, management, NCBI submissions, outreach & education — Eversole & Stein (IWGSC)

Germplasm Collection: Sourdille (INRAE)
- 8 landraces + Chinese Spring
- 4 landraces

9 accessions: Sequence & assembly — Corteva Agriscience, Genomics Technologies (Johnston, IA)

4 accessions: Sequence & assembly — Benhamed (U. Paris-Saclay); Poncet (INRAE)

13 accessions: mRNA Iso-Seq — Benhamed (U. Paris-Saclay)
Annotation — Choulet & Rimbert (INRAE)
IWGSC Data Repository — Alaux (INRAE)

Practical Haplotype Graph: 13 accessions + other published genomes Outreach & education — Buell (U. Georgia)

Aim 1: Generate platinum quality genome assemblies for 12 maximally diverse bread wheat landrace accessions and the community reference cv. Chinese Spring (CS)
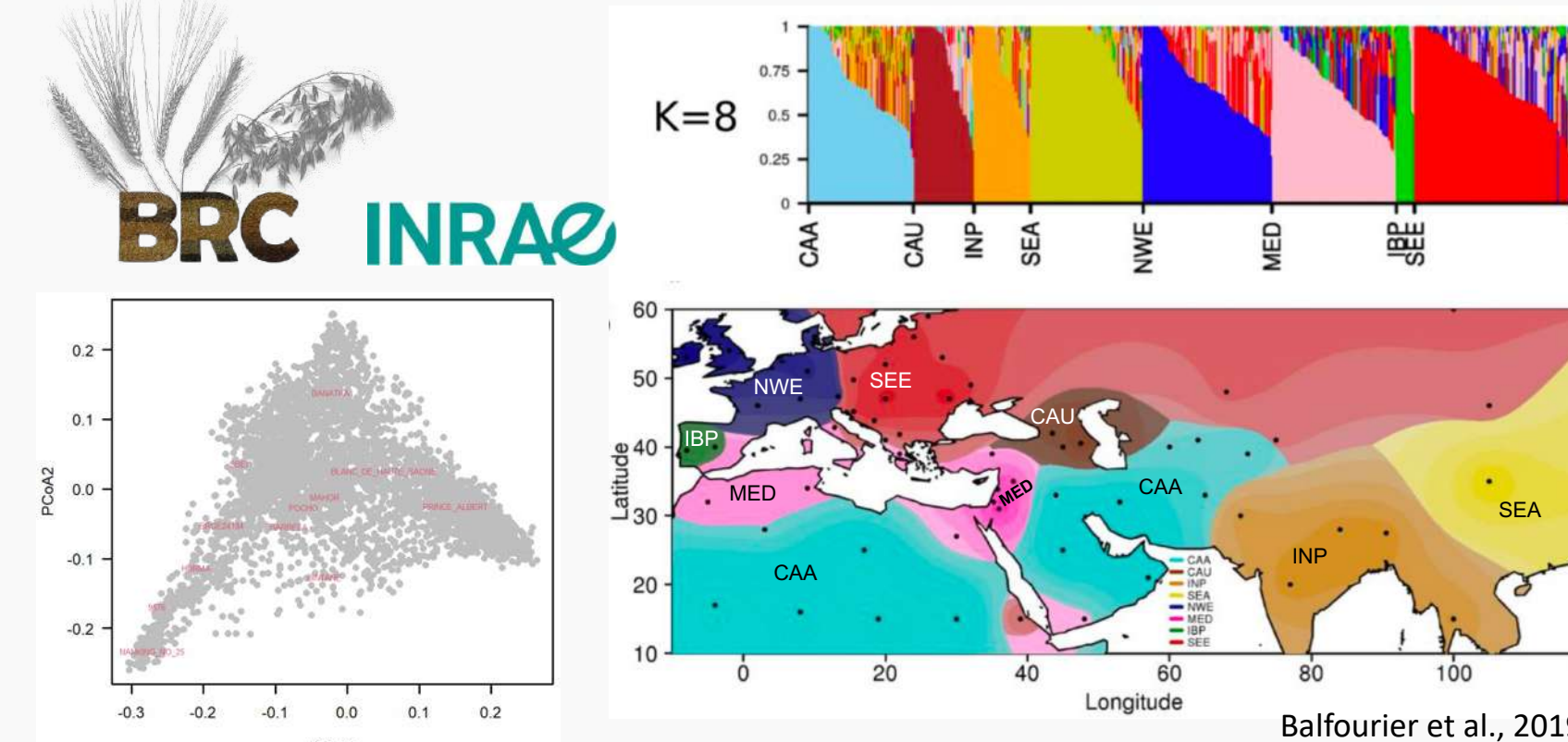
- 8 Phylogeographic haplogroups among 632 genotyped landraces = subpopulations corresponding to early regional growing centers.
- Genome sequence & assembly using long-reads, optical mapping, and conformational capture sequence.

Balfourier et al., 2019.

Table 1. Twelve accessions selected for sequencing in addition to cv. Chinese Spring.

Name	Haplogroup	Country	Habit	Assign-8†	Assign-12†	Accession‡
HURMA	CAA	Iraq	Spring	0.994	0.892	ERGE35883
ERGE24184	CAU	Palestine	Winter	1	1	ERGE24184
9676	INP	India	Spring	1	1	ERGE35970
NANKING_NO_25	SEA	China	Winter	1	1	ERGE5116
PRINCE_ALBERT	NWE	France	Winter	1	1	ERGE5858
MAHOR	MED	Tunisia	Spring	1	1	ERGE35861
BARBELA	IBP	Portugal	Spring	1	1	ERGE24043
BANATKA	SEE	Ukraine	Winter	1	1	
BINTANE*	NWE	Turkey	Spring	0.676	1	ERGE35937
BLANC_DE_HAUTE_SAONE*	NWE	France	Winter	0.506	1	ERGE7966
SBEI*	SEE	Morocco	Spring	0.714	1	ERGE35958
POCHO*	NA	Colombia	Spring	NA	NA	ERGE35904

†Assignment-to-group using 8-group or 12-group estimations; ‡Biological Resource Centre INRAE. *Genome sequence by collaborators in France.
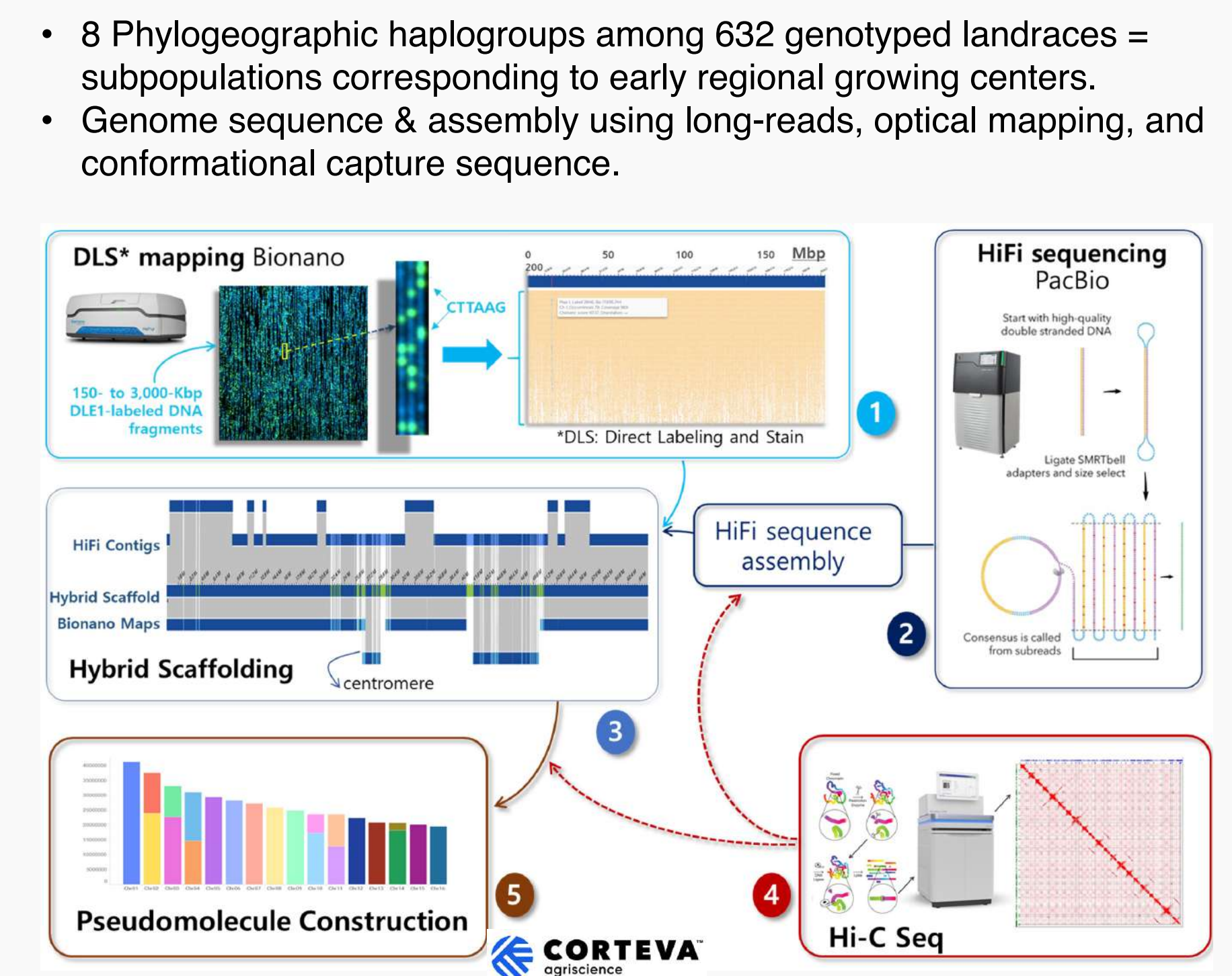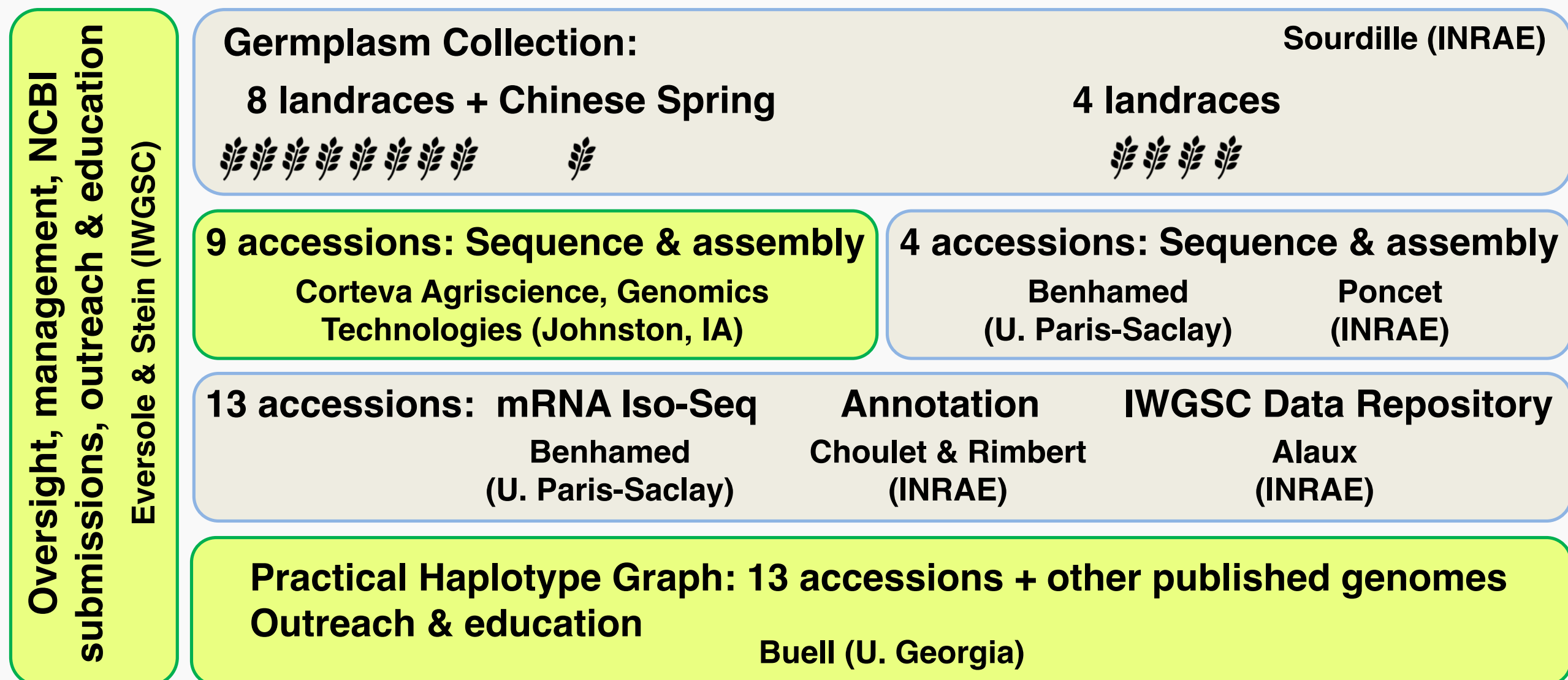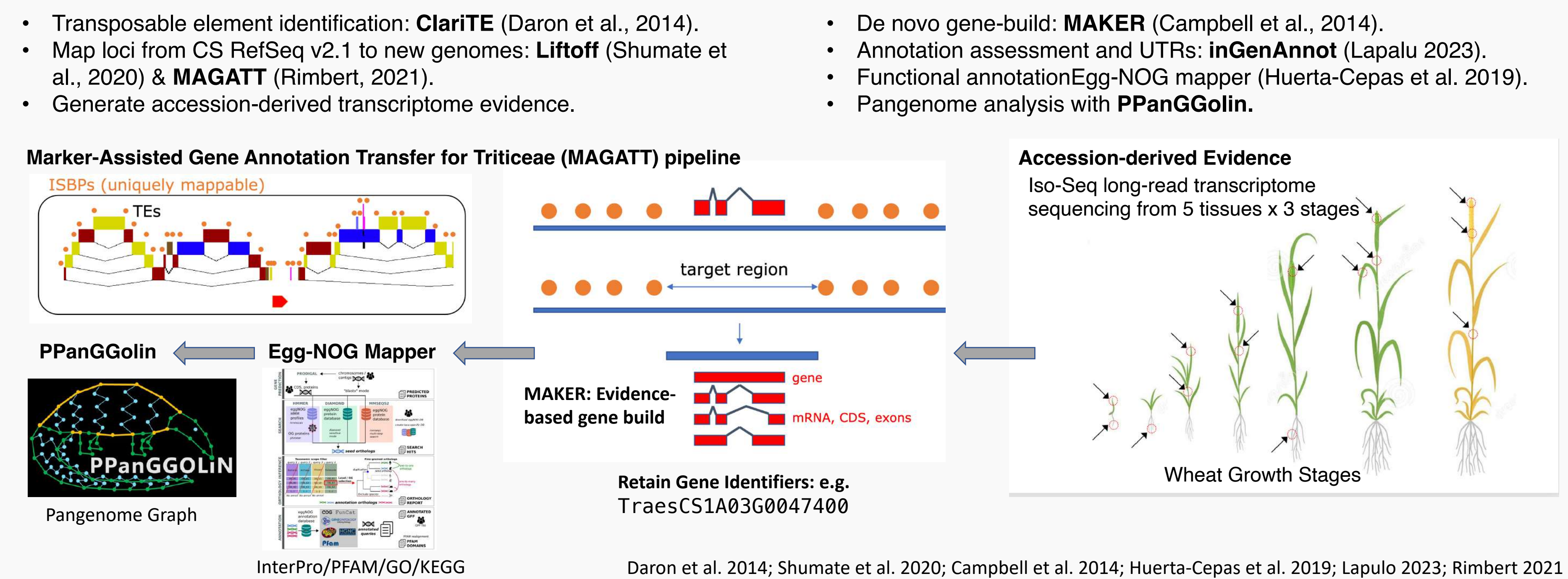
Aim 2: Establish transcriptome data and gene annotations for these accessions and update the IWGSC CS RefSeq annotation in a manner that sustains its role as a community-curated resource

- Transposable element identification: **ClariTE** (Daron et al., 2014).
- Map loci from CS RefSeq v2.1 to new genomes: **Liftoff** (Shumate et al., 2020) & **MAGATT** (Rimbert, 2021).
- Generate accession-derived transcriptome evidence.
- De novo gene-build: **MAKER** (Campbell et al., 2014).
- Annotation assessment and UTRs: **inGenAnnot** (Lapalu 2023).
- Functional annotation Egg-NOG mapper (Huerta-Cepas et al. 2019).
- Pangenome analysis with **PPanGGolin**.

Marker-Assisted Gene Annotation Transfer for Triticeae (MAGATT) pipeline

Accession-derived Evidence — Iso-Seq long-read transcriptome sequencing from 5 tissues x 3 stages

Wheat Growth Stages

MAKER: Evidence-based gene build. Retain Gene Identifiers: e.g. TraesCS1A03G0047400

InterPro/PFAM/GO/KEGG

Daron et al. 2014; Shumate et al. 2020; Campbell et al. 2014; Huerta-Cepas et al. 2019; Lapulo 2023; Rimbert 2021

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

Graph-based representation of genic and intergenic regions representing diversity across taxa.

Bradbury et al., 2022.

References:

Balfourier, F, Bouchet S, Robert, S, De Oliveira R, Rimbert H, Kitt J, Choulet F, International Wheat Genome Sequencing Consortium; BreedWheat Consortium; Paux, E. "Worldwide Phylogeography and History of Wheat Genetic Diversity". *Sci Adv* 2019, 5 (5), eaav0536.

Daron J, Glover N, Pingault L, Theil, S, Jamilloux, V, Paux, E, Barbe V, Mangenot, S, Alberti A, Wincker P, Quesneville H, Feuillet C, Choulet F. "Organization and Evolution of Transposable Elements along the Bread Wheat Chromosome 3B". *Genome Biol.* 2014, 15 (12), 546. https://doi.org/10.1186/s13059-014-0546-4.

Shumate, Alaina, Salzberg SL. 2020. "Liftoff: Accurate Mapping of Gene Annotations." Bioinformatics, December.

Bradbury PJ, Casstevens T, Jensen SE, Johnson, LC, Miller ZR, Monier B, Romay MC, Song B, BuckleR ES. "The Practical Haplotype Graph, a Platform for Storing and Using Pangenomes for Imputation". *Bioinformatics* 2022, 38 (15), 3698–3702.

Campbell MS, Law M, Holt C, Stein JC, et al. "MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations". Plant Physiol. 2014 Feb;164(2):513-24.

Huerta-Cepas, J, Szklarczyk, D, Heller, D, Hernández-Plaza, A, Forslund, SK., Cook, H., Mende, DR, et al. 2019. "eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses." Nucleic Acids Research 47 (D1).

Lapalu, Nicolas. 2023. "BIOGER / Ingenannot · GitLab." https://forgemia.inra.fr/bioger/ingenannot.

"PPanGGOLiN: Depicting Microbial Diversity via a Partitioned Pangenome Graph I PLOS Computational Biology." n.d. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1007732.

Rimbert, Hélène (INRAE/GDEC). 2021. "UMR GDEC / Magatt · GitLab." https://forgemia.inra.fr/umr-gdec/magatt.