Arima CiFi: A long read-based protocol for chromatin conformation capture for scaffolding plant genome assemblies

Meet us at booth #204!

Jayson Talag*, Gowdaman Vasudevan*, Chandler Sobel-Sorenson*, Branden Lau^, Greg Clarke*, Seunghee Lee*, Giovanni Melandri#, Alexander Bucksch#, Brian Kilburn†, Ibrahim Jivanjee†, Allyson Whittaker†, Rod A. Wing*, Dario Copetti*

*Arizona Genomics Institute, School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ USA; ^ Arizona Genetics Core, University of Arizona, Tucson, AZ USA; # School of Plant Sciences, University of Arizona, Tucson, AZ USA; †Arima Genomics, Carlsbad, CA, USA.

The advent of PacBio HiFi sequencing has considerably improved the quality of plant genome assemblies, reaching unprecedented completeness and base accuracy. However, chromosome-scale scaffolding, haplotype phasing, and 3D conformation are aspects that in most cases can't be addressed by HiFi reads alone.

Hi-C is the most common data type used to fulfill these needs; however, it suffers from the shortcomings associated with short reads – ambiguous mapping and low mappability in repeated regions.

Sequencing longer portions of the proximally ligated DNA will enable to map 3C data in repetitive and complex genomic regions.

The Dennis Lab at UC Davis has developed the CiFi method (McGinty *et al.*, 2025) on human and insect samples.

We adapted the CiFi protocol using the Arima CiFi kit to plant species of different size and complexity. The resulting matrices are comparable to the ones generated with Hi-C data. CiFi libraries can be sequenced with the HiFi library in the same SMRT cell.

Special thanks to:

ARIMA GENOMICS

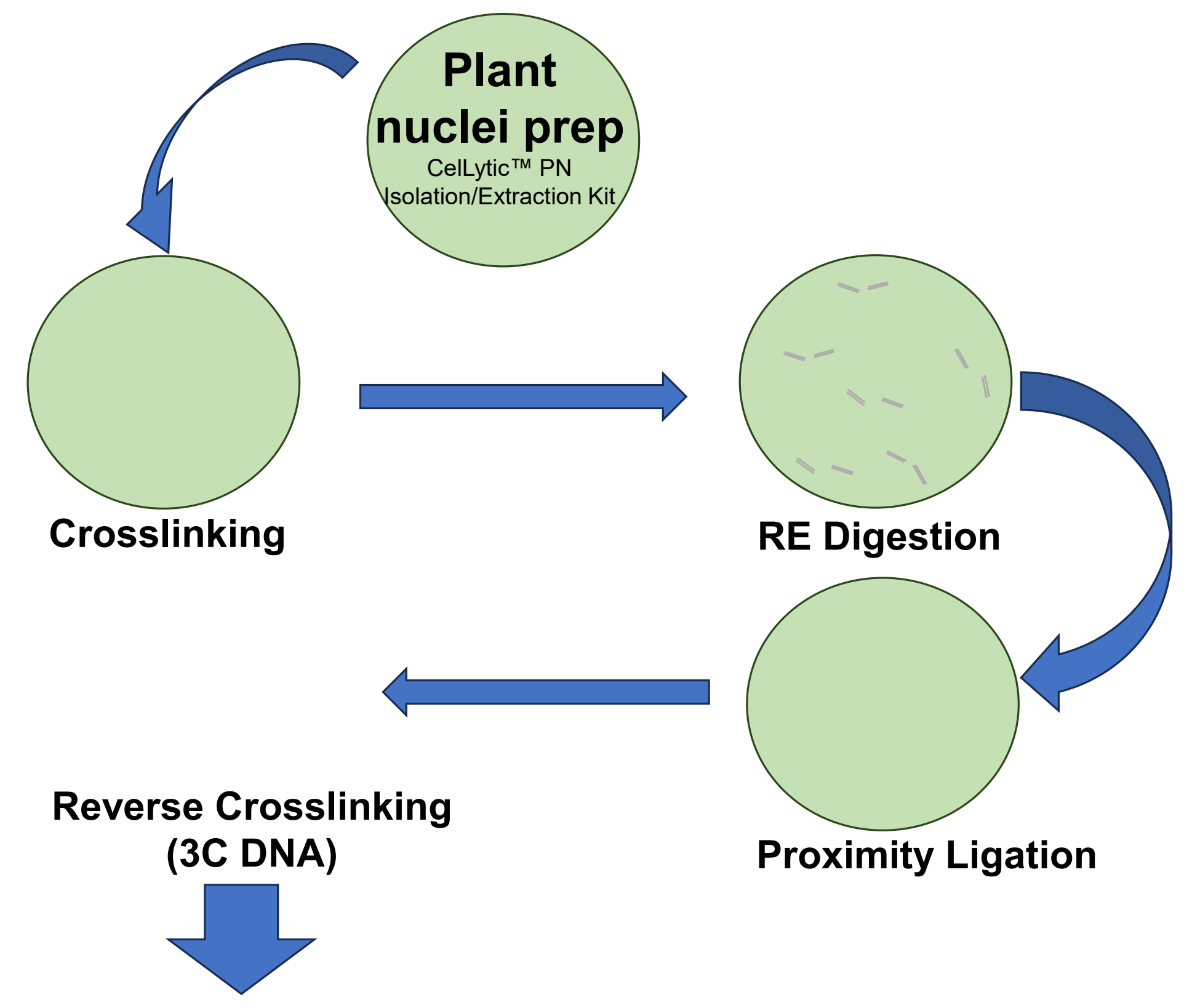
PacBio

dennis lab

HUMAN GENOMICS, EVOLUTION, DISEASE GENETICS

Nuclei extraction

Nuclei are extracted from leaves or other suitable plant organs.

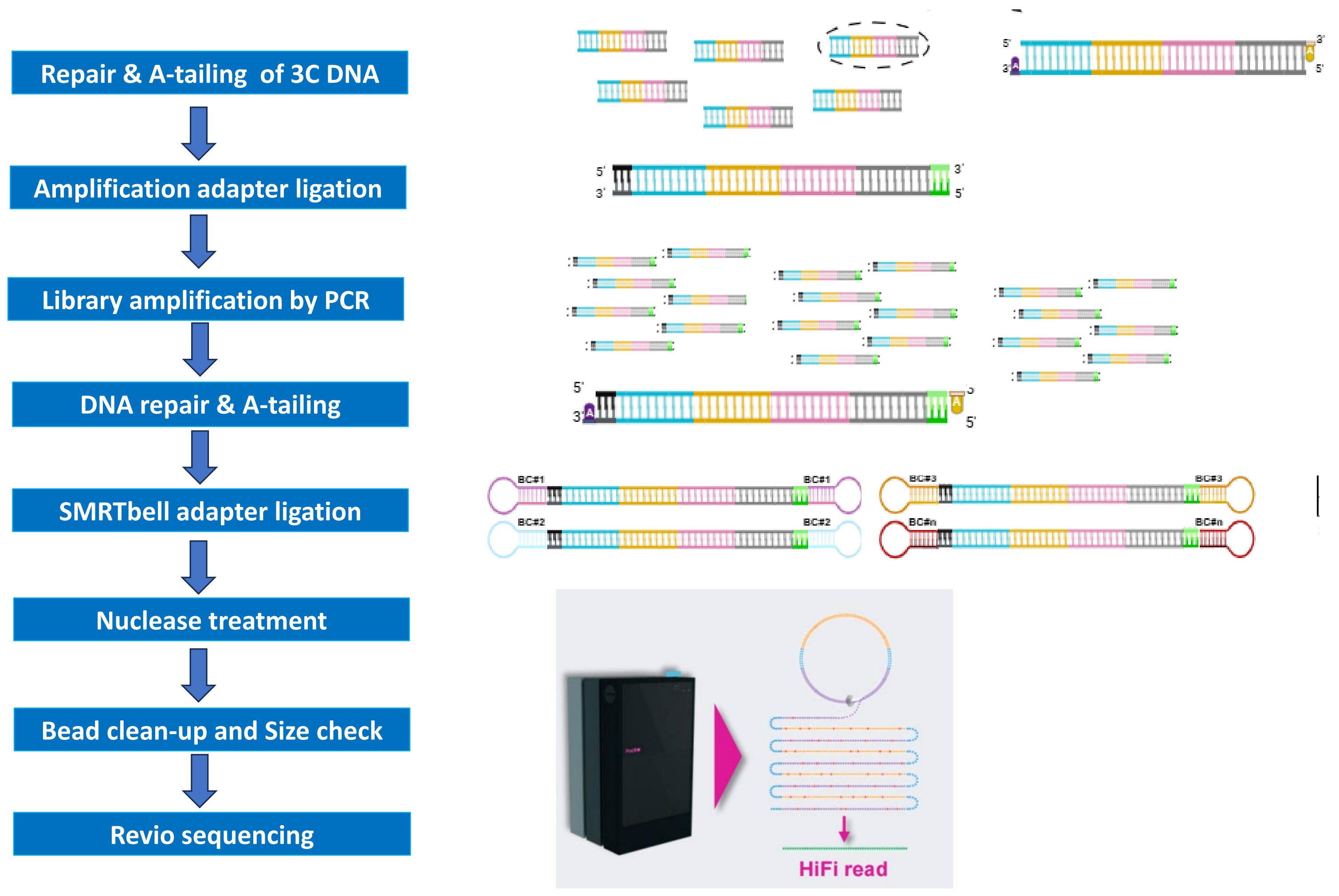


Arima CiFi (3C)

Nuclei undergo crosslinking, single-enzyme digestion, and proximity ligation to capture the spatial relationships of the DNA strands

PacBio Ampli-Fi sequencing

Concatenations of multiple crosslinking and proximally ligated events (2-8 kb) are amplified with PacBio's Ampli-Fi protocol. This yields sufficient DNA to be sequenced on the Revio platform, alone or in pools with other samples.


Repair & A-tailing of 3C DNA
Amplification adapter ligation
Library amplification by PCR
DNA repair & A-tailing
SMRTbell adapter ligation
Nuclease treatment
Bead clean-up and Size check
Revio sequencing

CiFi Computational Pipeline

mydennislab / CiFi

Ci-Fi data analysis and assembly scaffolding

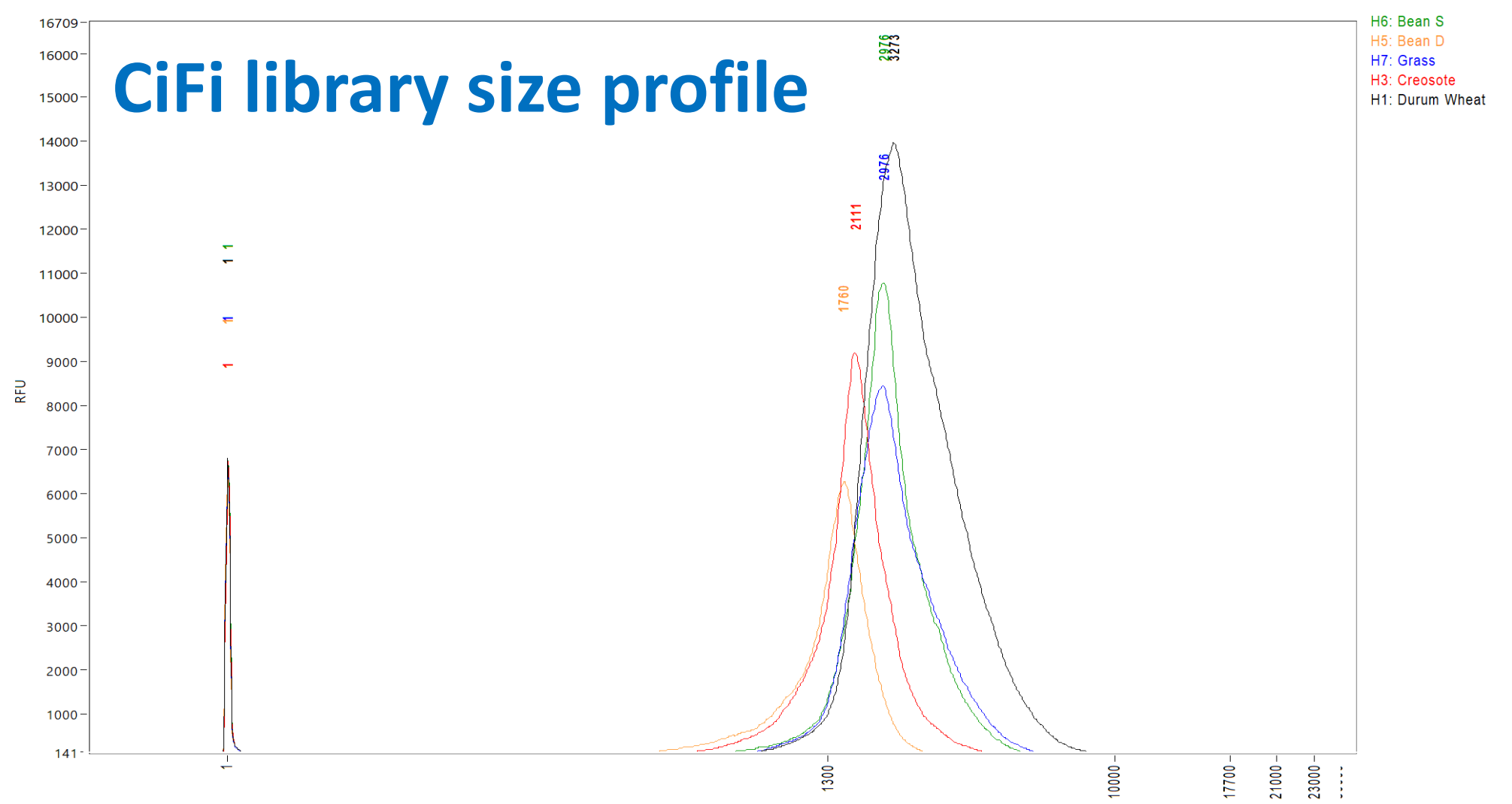
By mapping CiFi reads to each assembly, the individual segments are identified. Pairs of synthetic paired-end reads are generated from all the segments of a CiFi read. Then, CiFi synthetic read pairs are used in conventional Hi-C scaffolding software to generate a scaffolding matrix.

The Arizona Genomics Institute is a PacBio **PACBIO CERTIFIED SERVICE PROVIDER** **Request a quote:**

CiFi on five plant genomes

We adapted the CiFi protocol for plants, choosing species of different genome sizes, complexity, and ploidy that had HiFi assemblies available.

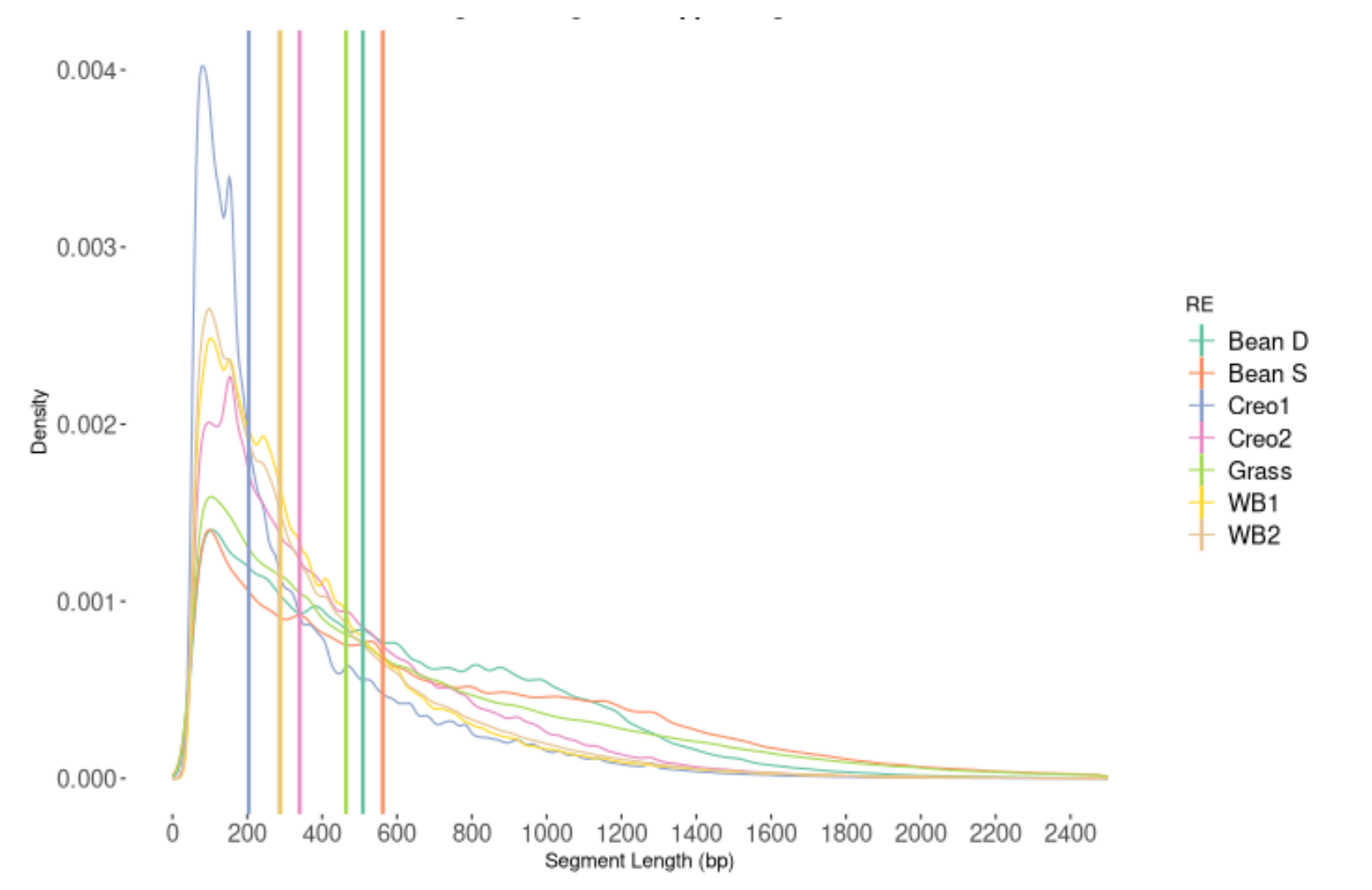
Sample	Bean S	Bean D	Grass	Creo	WB881
Common name	bean	bean	grass	creosote	durum wheat
Heterozygosity	homo	homo	hetero	hetero	homo
Ploidy	2n	2n	4n	4n?	4n
Total chr. number	22	22	36	52?	28
Haploid genome size (Mb)	750	750	1650	1450	5025
Assembly length (Mb)	596.9	592.2	3331.2	2883.0	10524.8
Number of sequences	367	221	1727	526	2513
N50 (Mb)	33.3	32.0	7.9	30.8	40.3
L50	7	7	110	37	48
N90 (Mb)	6.8	10.6	1.8	9.9	7.0
L90	20	22	464	101	300

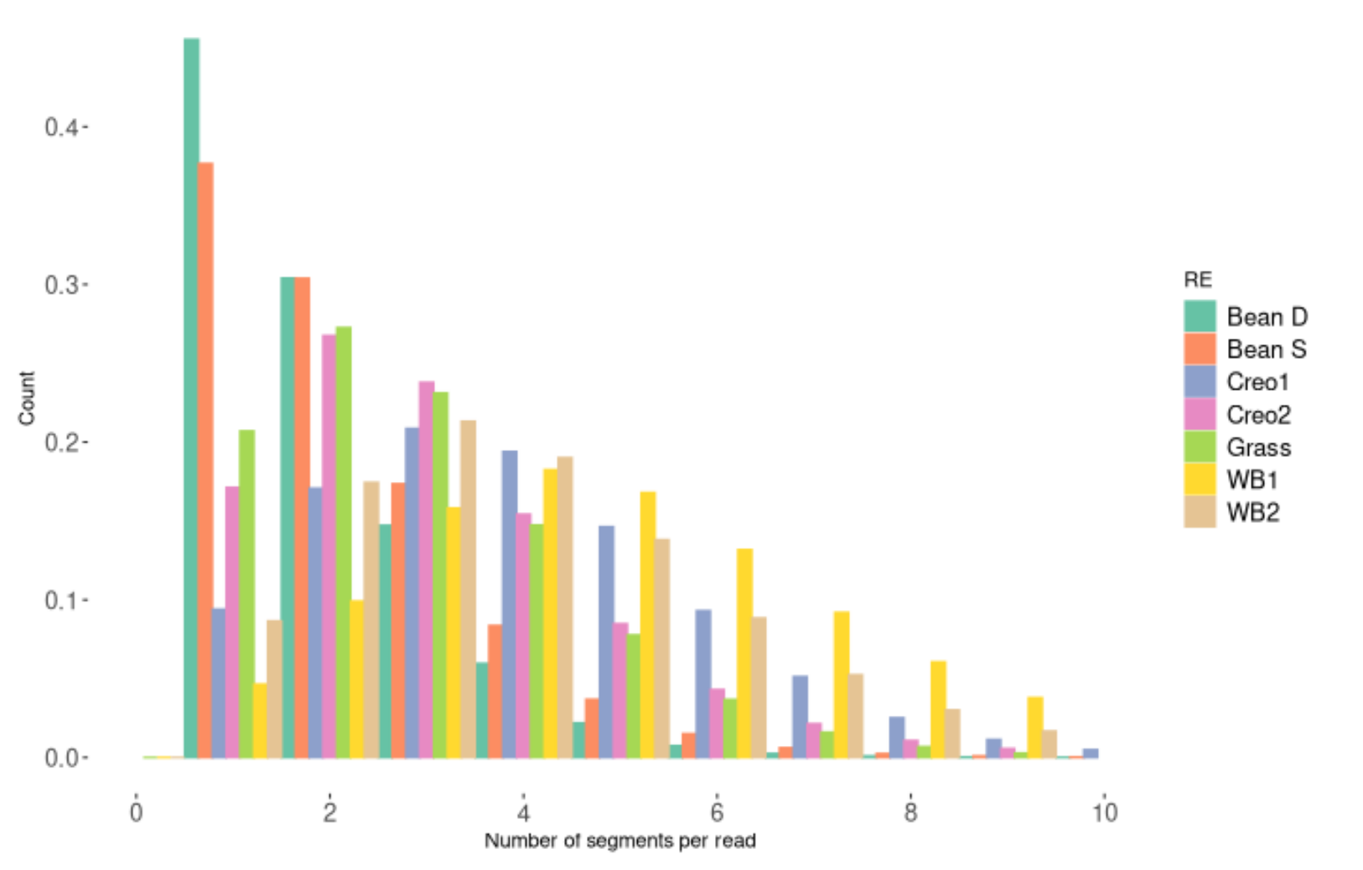


Plant CiFi libraries

The proximally ligated DNA produced concatemers between 1 to 10 kb in size. CiFi reads reflected the library size distribution.

CiFi read stats	Bean S	Bean D	Grass	Creo1	Creo2	WB1	WB2
Number of reads (M)	2.20	1.69	1.89	1.49	2.20	2.14	2.29
Avg read length (kb)	1.59	1.20	1.83	1.31	1.38	2.01	1.61
Total bases (Gb)	3.49	2.03	3.46	1.95	3.04	4.29	3.70
Avg segment length (bp)	712	597	631	326	440	388	392
Avg segments per read	2.20	1.93	2.86	3.88	3.08	5.06	4.02





Segment length and number

Different libraries displayed varying lengths and number of the re-ligated segments, a consequence of genomic features, enzyme used, and ligation conditions.

Hi-C contact matrix

The Juicer suite was used to process our CiFi data, generating a contact matrix equivalent to ones generated with short read-based Hi-C libraries. The presence of localized, strong signals off the diagonal provides the information to scaffold contigs.