Scalable Genotyping By Sequencing (GBS) Using Low-Pass Whole Genome Sequencing (lpWGS) and a One-Step, TnX™ Transposase-Based Library Prep Method for Genomic Selection

Michelle Rahardja, Sabina Gude , Rebecca Feeley, Jenna Couture, Stella Huang, Maura Costello, Yanyan Liu, Jack Leonard, Joe Mellor

seqWell, Inc. Beverly, MA USA

Introduction

Genomic prediction enables breeders to accelerate selection cycles by using genome-wide marker data to rapidly and cost-effectively assess genetic potential, but success depends on obtaining genotypes with sufficient density and accuracy at scale. Low-pass whole genome sequencing (lpWGS) with imputation has emerged as a flexible, discovery-enabled alternative to arrays, offering improved genome representation for diverse and structurally complex crop species. To support practical adoption, lpWGS workflows require highly uniform, scalable, and robust library preparation across large sample sets. Here, we evaluate AgriPrep™, a one-step, high-throughput library preparation method optimized for lpWGS, in soybean, rice, and barley. Using normalized three 6-plex libraries sequenced on the NovaSeq X Plus and downsampled to 1x for variant calling and imputation with Khufu Analysis Service (HudsonAlpha), we assess coverage uniformity and genotype accuracy to demonstrate AgriPrep's suitability for routine, cost-effective lpWGS in agricultural genomics.

AgriPrep Library Preparation Kit Workflow

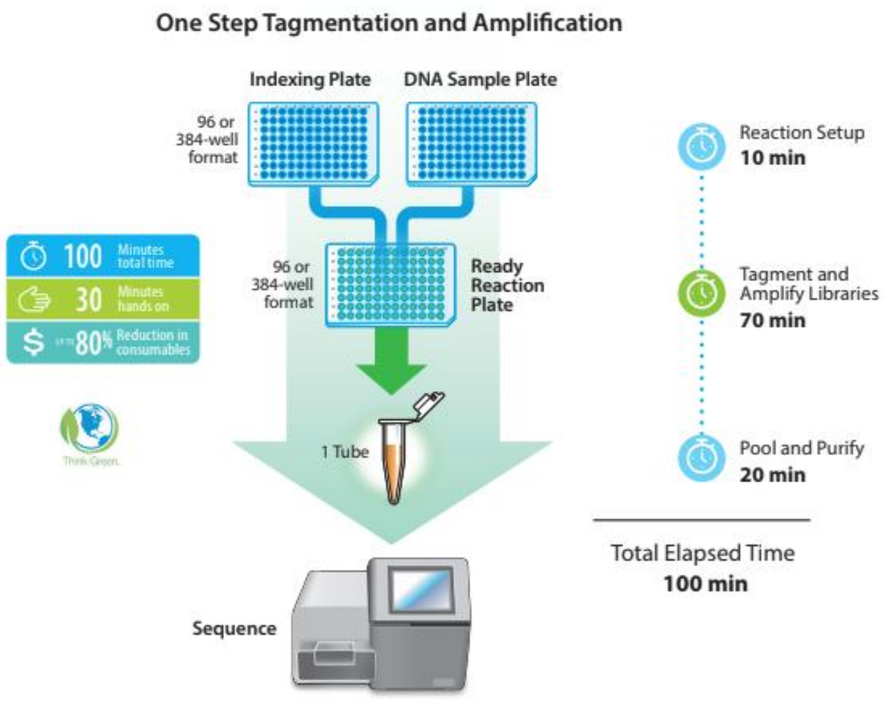

Figure 1. AgriPrep uses seqWell's high performance TnX transposase that was specifically engineered for NGS library preparation. The AgriPrep library prep kits utilize a proprietary mixture of enzymes to tag input DNA with indexed adapters and amplify libraries all in a single reaction. Additional index sets achieve multiplex levels >96. Samples are pooled volumetrically, purified, and converted into libraries to complete the 100-minute workflow, which includes 30 minutes of hands-on time. Built-in auto-normalization obviates the need to normalize sample input.

Methods

- The AgriPrep library prep was used to processed three different plant samples using 20, 50, and 100 ng input in duplicate (Table 1), creating three different pools of 6-plex.
- The libraries prepared were sequenced on an Illumina NovaSeq X Plus (300 cycles).
- Sequencing data were demultiplexed and aligned via Picard Tools. Variant calling analysis was done using the Khufu Analysis Service (HudsonAlpha, Huntsville, AL).

Table 1. Summary of plant samples assessed in the study.

Samples	Genome Size (Gb)	Description
Barley *(Hordeum vulgare)*	5.3	Large, repeat-rich genome
Soybean (*Glycine max)*	1.1	Medium, moderately repetitive genome
Rice *(Oryza sativa)*	0.4	Compact, low-repetitive genome

Coverage and Genotype Performance Metrics

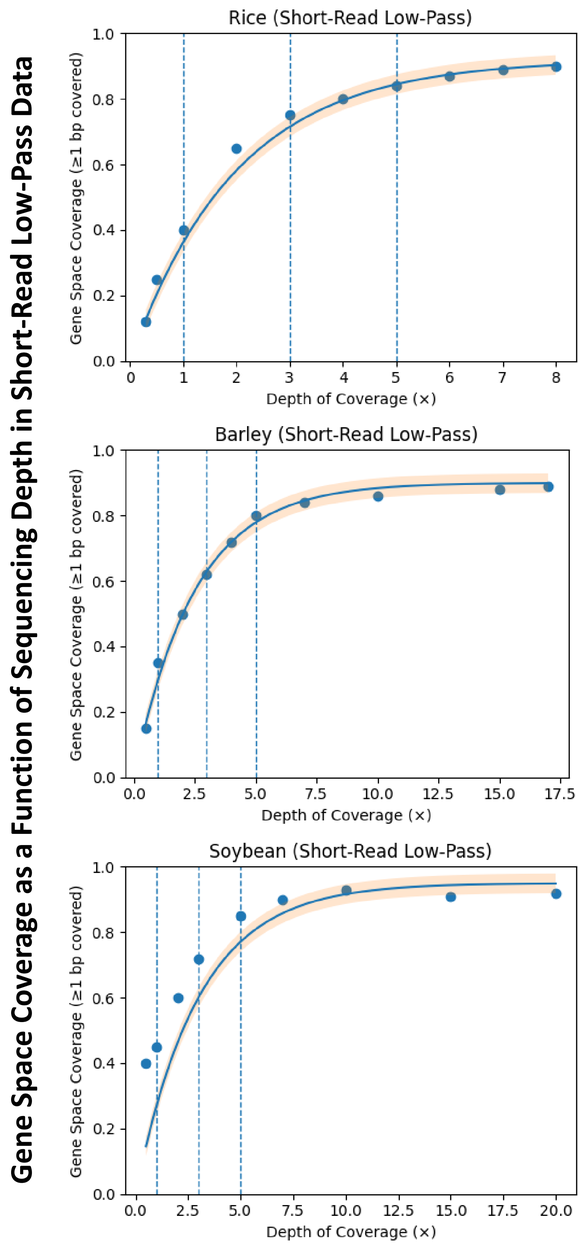

Figure 2. *Left:* **Gene space coverage as a function of sequencing depth across crop genomes using short-read low-pass sequencing.** Gene space coverage, defined as the fraction of annotated genes with ≥1 bp covered, increases rapidly at low sequencing depths in barley, soybean, and rice, with most gene space recovered by ~3-5x coverage and diminishing gains thereafter. Points represent individual samples, solid lines indicate fitted trends, and shaded regions denote confidence intervals. Vertical dashed lines mark representative low-pass coverage thresholds (1x, 3x, and 5x).

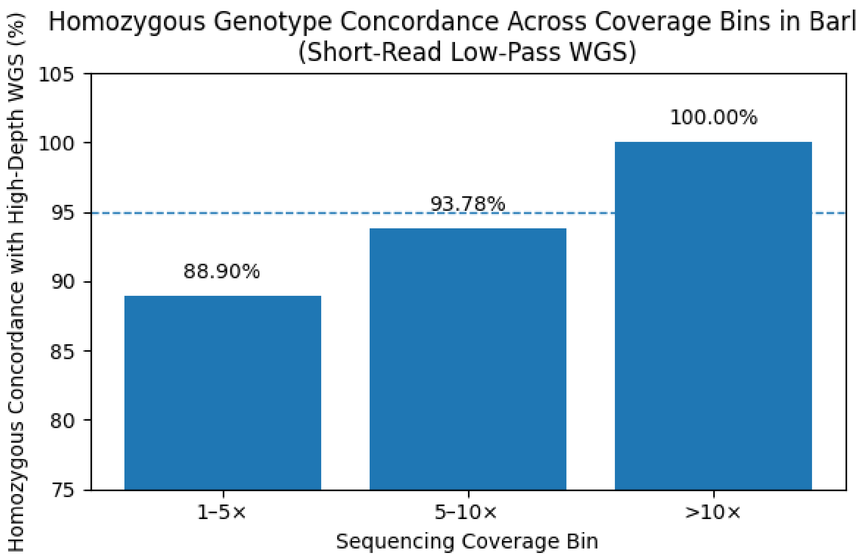

Figure 3. *Top:* **Homozygous genotype concordance with high-depth whole genome sequencing across sequencing coverage bins in barley.** Homozygous genotype calls derived from short-read low-pass whole genome sequencing were compared to high-depth WGS ground truth across three coverage ranges. Concordance increases from 88.90% at 1–5× coverage to 93.78% at 5–10× and reaches 100% at >10× coverage, indicating improved homozygous genotype accuracy with increasing effective sequencing depth in barley.

Sequencing Quality Metrics

Table 2. Sequencing metrics summary of three different pools of 6-plex prepared using AgriPrep on the NovaSeq X Plus post down-sampled to <1x.

Samples	Avg. Mean Coverage	Mean Insert Size (bp)	Duplication Rate
Barley *(Hordeum vulgare)*	0.54	459	1.7%
Soybean (*Glycine max)*	0.75	415	0.9%
Rice *(Oryza sativa)*	0.57	438	0.7%

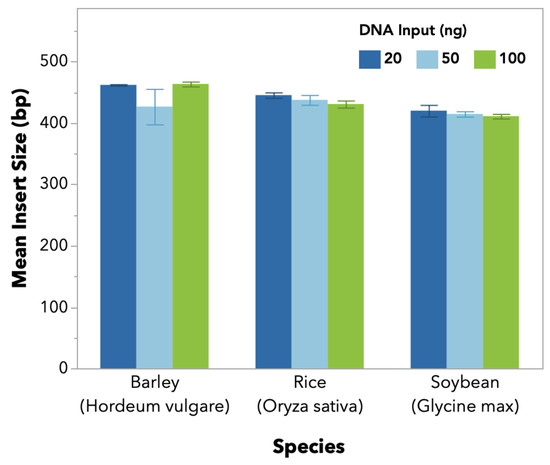

Figure 4. AgriPrep produces consistent insert sizes across species and DNA input amounts. Mean insert sizes for barley, rice, and soybean libraries prepared using AgriPrep at 20-100 ng DNA input show minimal variability, supporting robust high-throughput low-pass WGS library preparation. Error bars indicate standard deviation.

Summary and Conclusions

- AgriPrep enables robust, scalable short-read low-pass whole genome sequencing across diverse crop genomes, producing consistent insert size distributions (Figure 4) across species, sample types, and input amounts to support uniform genome-wide sampling.
- Gene space coverage increases rapidly at modest sequencing depths, with most annotated genes represented at approximately 3-5x coverage, demonstrating efficient genome sampling using low-pass sequencing (Figure 2).
- Combined with downstream imputation, AgriPrep enabled low-pass WGS provides a cost-effective, high-throughput genotyping solution suitable for genomic selection and agricultural breeding programs.