# Scalable Sample Extraction and Multiplexed Library Preparation for High-Throughput Crop Genotyping Via Low-Coverage Whole Genome Sequencing

Michelle Rahardja[1], Jessica Smith[1], Ariele Hanek[1], Joe Mellor[1], Tobias Schughart[2], Maximilian Weiter[2] and Jesse L. Hoff[3]

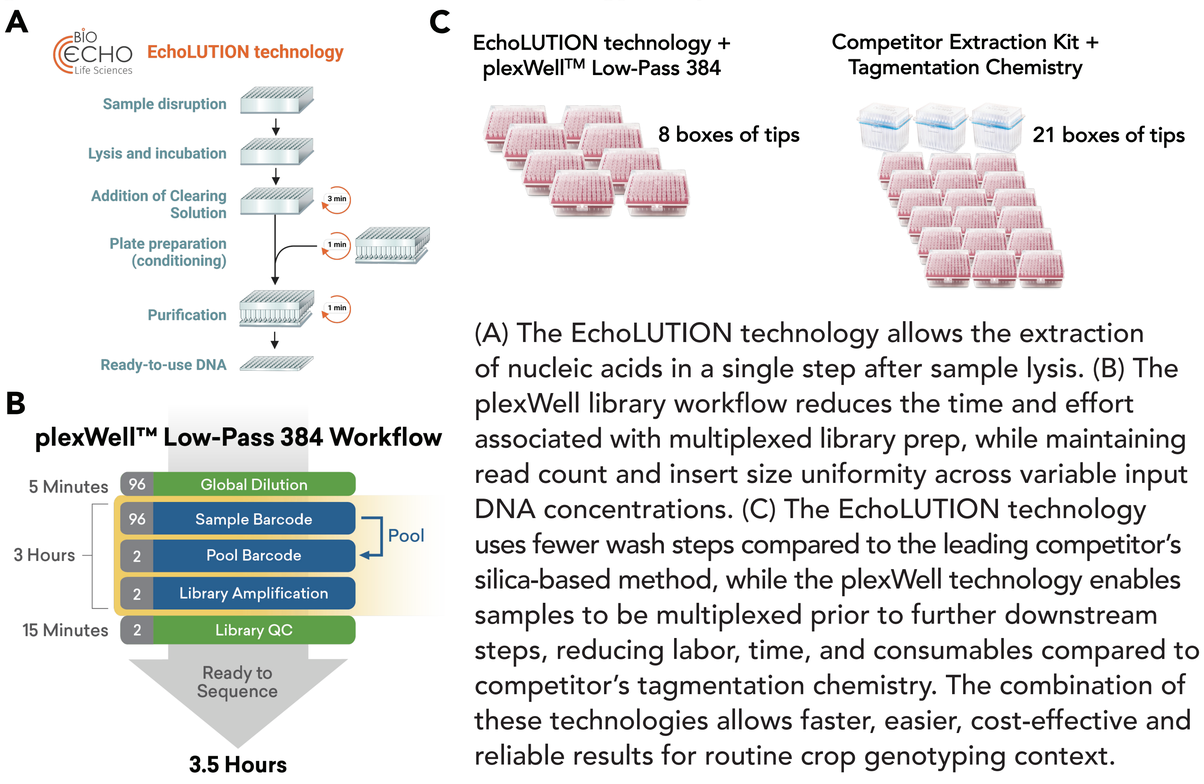
[1]seqWell, Inc. Beverly, MA USA, [2]BioEcho Life Sciences GmbH, Cologne, Germany, [3]Gencove, Inc. Los Angeles, CA USA

## Introduction

The reduced cost and improved scale of DNA sequencing has allowed a variety of genotyping assays to be approached with sequencing-based alternatives over legacy approaches such as microarrays. One key remaining bottleneck to unlock the efficient use of sequencing instruments for high-throughput genotyping is the creation of scalable workflows and chemistry for sample extraction and library preparation. Here, we demonstrate the robust combination of the EchoLUTION platform (BioEcho Life Sciences) and the plexWell™ Low-Pass multiplexed NGS technology (seqWell) across multiple plant species of different genome sizes. The EchoLUTION Plant DNA Kits enable the extraction of genomic DNA from different plant species and tissues, while plexWell Low-Pass Library Preparation Kits create normalized multiplexed libraries without the need for time-consuming adjustment of input DNA, significantly simplifying the complex task of high-level multiplexing. Our results show the utility of combining EchoLUTION and plexWell multiplexed NGS technologies to enable faster, easier, cost-effective, and reliable results for routine crop genotyping applications.

## EchoLUTION and plexWell™ Low-Pass 384 workflows

**Fig. 1: Combination of the EchoLUTION technology and plexWell Low-Pass 384 Workflow**



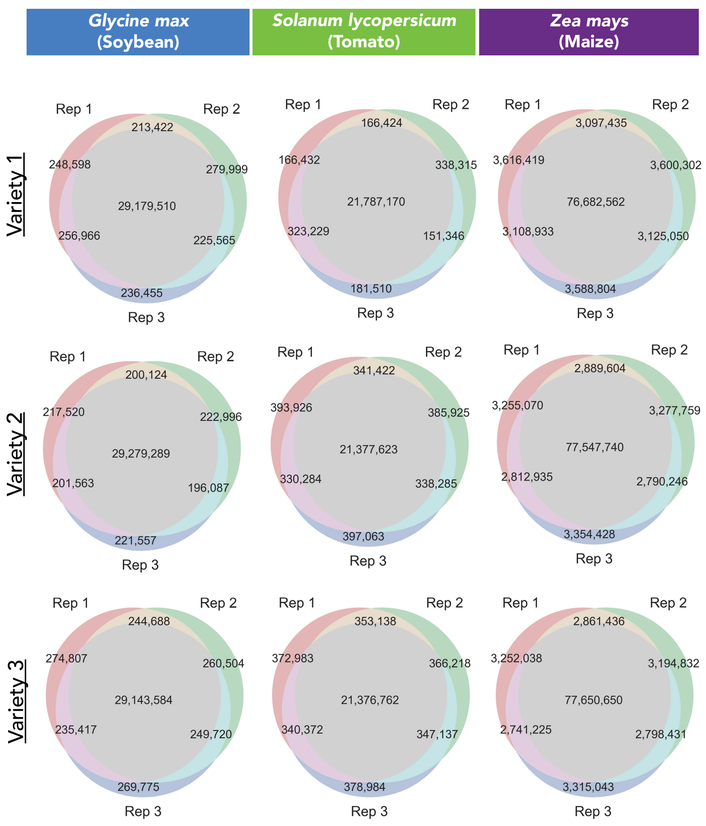

(A) The EchoLUTION technology allows the extraction of nucleic acids in a single step after sample lysis. (B) The plexWell library workflow reduces the time and effort associated with multiplexed library prep, while maintaining read count and insert size uniformity across variable input DNA concentrations. (C) The EchoLUTION technology uses fewer wash steps compared to the leading competitor's silica-based method, while the plexWell technology enables samples to be multiplexed prior to further downstream steps, reducing labor, time, and consumables compared to competitor's tagmentation chemistry. The combination of these technologies allows faster, easier, cost-effective and reliable results for routine crop genotyping context.

## Methods

- A collection of plant seeds encompassing 3 varieties of each of *Glycine max* (soybean), *Solanum lycopersicum* (tomato), and *Zea mays* (maize) was extracted in triplicate to produce 27 purified genomic DNA using the EchoLUTION Plant DNA kit from BioEcho Life Sciences.
- All samples within its species were globally diluted to an average of 1.7 ng/µl. Each sample was processed in duplicate with the plexWell Low-Pass library preparation kit, resulting a normalized multiplexed library pool.
- The library was sequenced on NextSeq2000 (2x150) and down-sampled to 2X coverage prior to Gencove's imputation analysis.

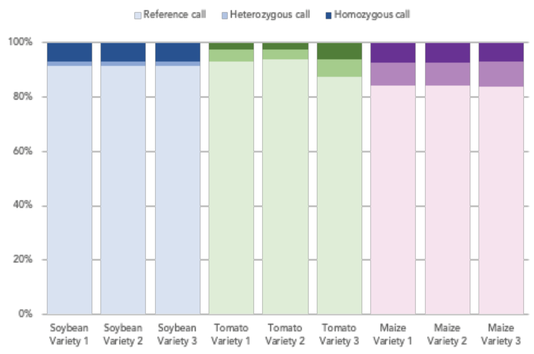
## Consistent Imputation Results

**Fig. 2: High SNP concordance across technical replicates for all species and varieties**



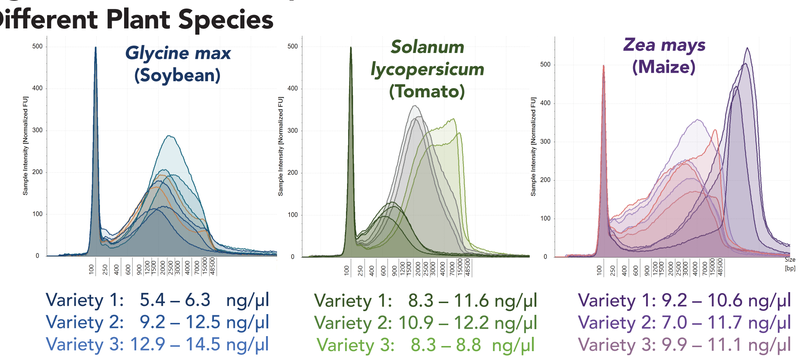

Each triplicate set for each variety of *Glycine max* (soybean), *Solanum lycopersicum* (tomato), and *Zea mays* (maize) were down- sampled to 2X coverage and analyzed using Gencove's imputation pipeline.>98% (soybean), >97% (tomato), and >91% (maize), of SNPs were called with high confidence. Figure 2 demonstrate the high concordance between technical replicates of soybean (95%), tomato (92%) and maize (80%). Reference, heterozygous, and homozygous calls are represented in Figure 3.

**Fig. 3: Percentage of SNPs called as reference, heterozygous or homozygous by plant and variety**




## Robust Sequencing Performance

**Fig. 4: Genomic DNA Tapescreen of Extracted DNA from Three Different Plant Species**




Variety 1: 5.4 – 6.3 ng/µl
Variety 2: 9.2 – 12.5 ng/µl
Variety 3: 12.9 – 14.5 ng/µl

Variety 1: 8.3 – 11.6 ng/µl
Variety 2: 10.9 – 12.2 ng/µl
Variety 3: 8.3 – 8.8 ng/µl

Variety 1: 9.2 – 10.6 ng/µl
Variety 2: 7.0 – 11.7 ng/µl
Variety 3: 9.9 – 11.1 ng/µl

**Fig. 5: Insert Size Uniformity of 54-plex Low-Pass WGS Multiplexed Library of Three Different Plant Species**




**Table 2: Characteristic plexWell Low-Pass 384 Library Performance**

| | *Glycine max* (Soybean) | *Solanum lycopersicum* (Tomato) | *Zea mays* (Maize) |
|---|---|---|---|
| % Reads Aligned | 98.7% | 98.4% | 99.1% |
| # of Reads generated | 374M | 443M | 531M |
| Read Count CV | 28.6% | 48.3% | 18.1% |
| Average Median Insert | 302 | 290 | 310 |

The EchoLUTION Plant DNA Kit extracts high-quality DNA (Fig. 4) consistently across multiple plant species of different genome sizes. The plexWell Low-Pass 384 Library Preparation Kit gives reproducible library characteristics (Table 2 and Fig. 5). All replicates within plant species analysis of read count and insert size are consistent and robust. Inserts size of 300-350 generate optimal clusters on Illumina Sequencers thus maximizing the unique data generated from 2x150 paired-end sequencing.

## Summary and Conclusions

- The robust combination of EchoLUTION and plexWell Low-Pass multiplexed NGS technologies provides a streamlined, scalable, and reproducible solution for routine crop genotyping.
- The imputation data at 2X coverage is consistent between the replicates of 3 different varieties in 3 different plant species of different genome sizes with an average concordance of 90% demonstrating wide ranging utility to AgBio high throughput sequencing applications.