APPLICATION NOTE

Efficient low-pass whole genome sequencing on the Ultima Genomics UG 100 sequencing platform.

ULTIMA GENOMICS

CURIO

Introduction

Low-pass whole genome sequencing (LP-WGS) provides a cost-effective, powerful and scalable approach to capturing genome-wide variation at coverage depths typically ranging from 0.1x to 10x. When combined with genotype imputation, LP-WGS enables accurate inference for unobserved and low-confidence variants and has emerged as a preferred genotyping strategy in crop and livestock breeding programs. Unlike fixed-content SNP arrays, LP-WGS offers an unbiased and flexible view of the genome, allowing researchers to detect both common and rare variants, structural rearrangements, and species-specific polymorphisms. This comprehensive variant detection accelerates genomic selection, enhances predictive accuracy in breeding models, and supports the discovery of novel loci linked to productivity, disease resistance, and environmental adaptation.

Despite its advantages, the adoption of LP-WGS at population scale has been constrained by the computational demands and data-handling complexity of imputation workflows, particularly when applied to populations numbering in the thousands. Platforms such as CURIO™ help overcome these challenges by enabling scalable, cloud-based analysis and imputation.

Sequencing platforms that deliver high-throughput, consistent coverage, and low per-sample cost are therefore critical to enable large-scale implementation in applied genomics and breeding pipelines.

The UG 100® sequencing platform from Ultima Genomics addresses these challenges through an innovative open-wafer, continuous-flow architecture that departs from conventional flow-cell-based designs. This approach allows highly parallel sequencing reactions across an unrestricted surface, minimizing reagent consumption and enabling uniform data generation at exceptional throughput. A single wafer typically yields 10-12 billion reads per run, providing cost-efficient coverage across large sample cohorts. The system's highly uniform coverage and low error rates support accurate variant calling across diverse genome types, making it particularly well suited for population-scale agrigenomic and breeding applications where throughput, consistency, and data quality are critical.

Methods

Sample preparation was performed at Texas A&M AgriLife (TX, USA). Six rice samples from clonal plants corresponding to the Presidio cultivar (tropical japonica) were homogenized mechanically using the Omni Bead Ruptor™ 96 bead mill homogenizer (Revvity, 27-0001), followed by DNA isolation using the chemagic™ 360 instrument (Revvity) with the chemagic DNA Plant 100 mg Kit (Revvity, CMG-795). Libraries were prepared using the NEXTFLEX™ HT Agrigenomics Low-Pass WGS Kit (Revvity) according to the manufacturer's instructions, and shipped to Ultima Genomics, Inc (CA, USA) for conversion and sequencing.

10 ng of input library were PCR amplified for 7 cycles using Ultima Genomics conversion primers[1]. After PCR, converted libraries were purified using SPRI beads and the final concentration was calculated using DNA concentration measured by Qubit® fluorometer (Thermo Fisher) and average library size as measured by Agilent® TapeStation® instrument. Libraries were sequenced on the UG 100 sequencer. CRAM files were uploaded to the CURIO™ platform for QC, alignment, and imputation analysis, using the Global *Oryza sativa* Reference Panel (GORP)[2].

This panel is a high-density resource that captures extensive allelic and haplotype diversity across *O. sativa* subpopulations. Its broad representation ensures high imputation accuracy across breeding materials, supporting genome-wide association and genomic prediction studies.

Low-coverage alignment data for imputation were created by downsampling the 80X to different sequencing depth thresholds ranging from 0.01X to 4.0X. The GLIMPSE2[3] imputation algorithm was applied to the downsampled alignments. The imputation analysis compared the LP-WGS data to the GORP panel to confidently infer genotypes using statistical algorithms that exploit linkage disequilibrium patterns. To evaluate the accuracy of imputation, we performed two non-reference concordance analyses: 1) "raw" concordance of all non-reference sites called in the 80X data and imputed in the low-coverage data for corresponding samples, and 2) "filtered" concordance of only high quality sites in the 80X data (SNVs with a depth [DP] ≥ 10 reads and a genotype quality Phred score [GQ] ≥ 30) and imputed data (genotype probability [GP] ≥ 95%).

Results

Data alignment

The CURIO™ platform analysed the 6 CRAM files (105 Gbp per sample on average) following the recommendations of Ultima Genomics. Time for processing was ~ 2 h for the full set of 6 samples.

We processed ~ 170 M reads per sample that were aligned using Ultima Aligner (UA) version 3.0.2. This corresponded to a depth of ~ 80x and 95.9% of covered positions in the rice exome. Coverage of chromosomes 1 to 12 is very similar (Figure 1).

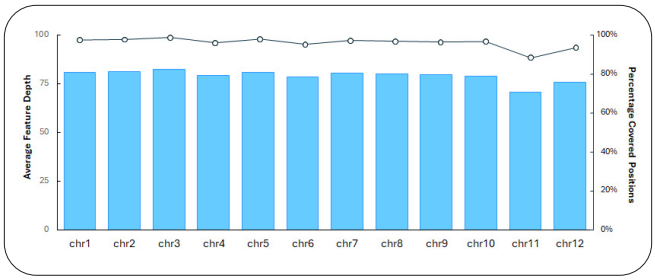

Figure 1: Average depth per chromosome for sample r1-2, downsampled to ~ 80x. A comparable percentage of covered annotated gene positions was observed across chromosome 1-12. Coverage analysis was restricted to exonic regions to minimize biases introduced by low-complexity or repeat-rich sequences present in non-coding DNA.

Variant analysis

Single nucleotide variants (SNVs) from homozygous or heterozygous positions with depth ≥ 2 were reported. Positions with indels were excluded. Source of truth (high coverage) was called using Freebayes variant caller. This caller was used as it is computationally efficient and can focus on a subset of the data, such as the GORP panel. More than 5 M variants were called for the 6 samples, and of these, approximately 680k were not homozygous reference genotypes, and therefore subsequently used for concordance analysis.

Imputation analysis

Concordance declines with lower coverage depth as expected, and in the case of unfiltered genotypes it went from 96.6% at 4x to 92.2% at 0.01x (Figure 2A). Filtered results decline less dramatically with decreasing sequencing depth, and generally achieved concordance values around 98%. Notably, at lower sequence depths, quality filtering of the imputed data results in less genotypes that pass the quality thresholds, though high concordance is still obtained (Figure 2B).

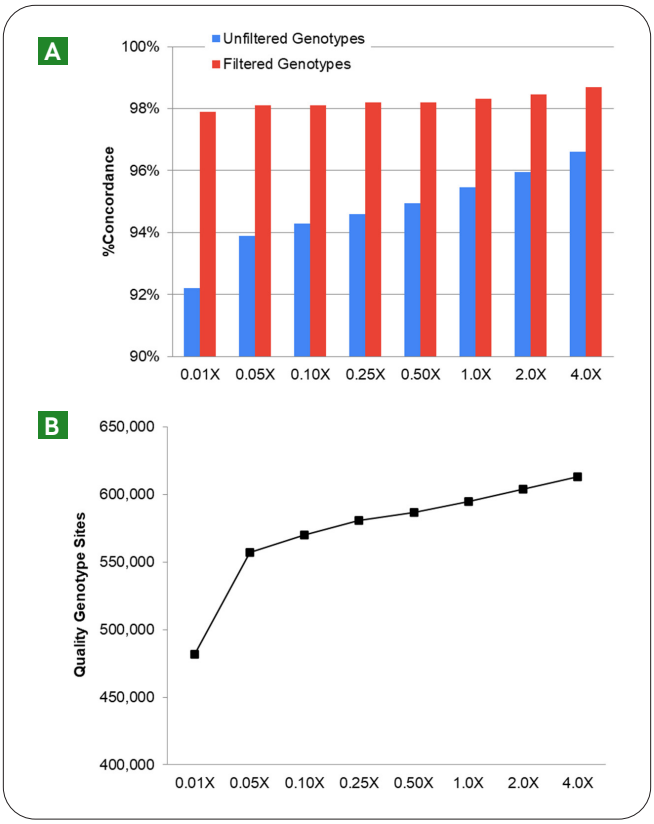

Figure 2: **A**. Percentage of concordance between imputed genotypes and source of truth data, across different coverage depths. Unfiltered and filtered genotypes are presented. **B**. Number of imputed genotype sites with GP ≥ 95% and validation sites with DP ≥ 10 reads and GQ ≥ 30 at different coverage depths. Results for A and B correspond to the average of the 6 samples.

To have a visual representation of the imputation, we took the data of sample r1-2 and inspected a random region of 3kb in chromosome 1 (chr1:2,661,956-2,664,955). CURIO™ allows comparison of the original data and the imputed results at different coverage depths. This shows that most of the variants called after imputation are identical to the variants observed at 80x (Figure 3).

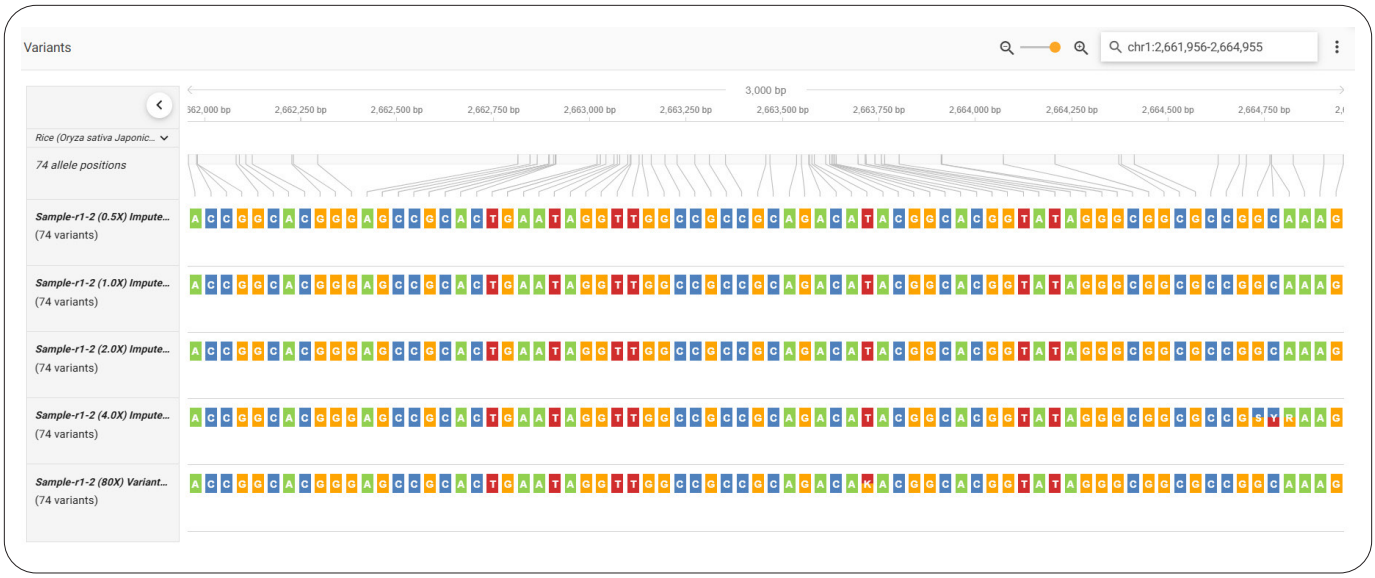

Figure 3: Window of 3kb in chr1 of sample r1-2 containing 74 observed variants at 80x. All of them variants are called in the imputed data at different sequencing depths. In this region, 100% of the imputed variants are concordant with observed data.

The genome size of *Oryza sativa* is approximately 390 Mb for its diploid form, the smallest among the domesticated cereals. With this size, only 6 million reads per sample are required to obtain 4x coverage depth, which means that ~2,500 samples could be processed in a single run with the UG 100® sequencing platform, and up to 20,000 at 0.5x coverage depth.

Conclusion

Sequencing performed on the UG 100® sequencing platform provides highly uniform coverage and consistent data quality, enhancing the scalability and cost-effectiveness of low-pass whole genome sequencing workflows. The combination of this technology with the NEXTFLEX HT Agrigenomics Low-Pass WGS Kit with the CURIO™ platform (curiogenomics.com) offers a fast and robust solution for genotyping any plant or animal species, eliminating the need for extensive bioinformatics expertise or computational infrastructure. This streamlined approach simplifies and democratizes high-throughput plant and animal studies from start to finish. Imputation concordance between data at 0.5x and 80x is very high for the samples, supporting the convenience of this approach even for very low coverage applications.

References

1. Solaris Free & Solaris Flex | Flexible Library Prep Workflows | Ultima Genomics
2. Wang, D.R., et al. (2018). An imputation platform to enhance integration of rice genetic resources. *Nat Commun* **9**, 3519. doi:10.1038/s41467-018-05538-1.
3. Rubinacci, S., et al. (2023). Imputation of low-coverage sequencing data from 150,119 UK Biobank genomes. *Nat Genet* **55**, 1088–1090. doi:10.1038/s41588-023-01438-3.