Leveraging hybridization capture for cost-effective and efficient large-scale genomic analysis

Targeted next-generation sequencing outperforms traditional whole-genome approaches in wheat genome case study

Executive Summary

Sequencing large and complex genomes, such as hexaploid wheat with its ~17 Gbp genome, using traditional whole genome NGS is costly. Targeted sequencing via the hybridization capture technique provides a focused approach, allowing researchers to sequence specific genes or regions of interest, dramatically reducing sequencing volume and computational burden.

In a case study targeting >200 Mbp of agronomically important wheat genes, hybridization capture reduced sequencing costs by 88%, lowered the data analysis computation burden, and provided high-depth coverage for precise variant detection. These efficiencies can accelerate variant discovery for agrigenomic crop improvement applications while optimizing project resource allocation.

Introduction

Background

Wheat has a large and complex genome, including multiple polyploid forms, presenting challenges for implementing next-generation sequencing (NGS) for large-scale, genome-wide discovery and screening projects. Performing genome-wide variant discovery with traditional methods necessitates sequencing the entire genome to the desired coverage depth, including wasting sequencing reads on non-informative and/or repetitive regions. Costs for this approach can easily exceed $1,700 per sample due to high requisite sequencing read depths for novel variant discovery, and the subsequent computational resources needed to process and analyze the large datasets.

Problem Statement

Traditional methods of conducting genome-wide variant discovery, such as whole genome sequencing, incur:

- High sequencing costs per sample
- Large data storage and processing requirements
- Reduced turnaround times that limit the speed of research and breeding programs

Solution Overview

Hybridization capture focuses sequencing on predefined genomic regions, allowing researchers to obtain high-quality data at a fraction of the cost and time of whole-genome approaches.

Hybridization Capture Technology

Principles

- Custom probes are designed to “capture” specific genomic regions of interest.
- Captured DNA is enriched, sequenced, and analyzed, providing deep coverage where it matters most.

Advantages Over Traditional NGS for Variant Discovery

Feature	Whole Genome Sequencing	Targeted Sequencing with Hybridization Capture	Improvements
Sequencing Volume	Dictated by genome size + desired read coverage depth *(e.g. minimum 17 Gbp for 1×, or ~1700 Gbp for 100× coverage)*	Scales with desired target size, coverage depth, and assay efficiency.	Fewer sequencing runs Faster data processing analysis
Cost per Sample	High; driven by sequencing and data processing	Minimized; efficient sequencing and data processing	Major cost reduction, depending on project type

---

Case Study: Wheat Exome

Objective

Analyze coding genes for diverse wheat cultivars using targeted NGS, enabling variant discovery for agrigenomic crop improvement applications including disease resistance, yield, and stress tolerance.

Methodology

- **Probe Design:** Custom 120 bp probes targeting coding genes based on hexaploid wheat reference genome (~200 Mbp total target space).
- **Sample Preparation:** DNA extraction from 24 hexaploid wheat specimens, library preparation with hybridization capture.
- **Sequencing:** Illumina NextSeq, paired-end 150 bp reads.
- **Data Analysis:** Alignment to reference genome, variant calling, and coverage assessment.

Results

Metric	Whole Genome Sequencing	Hybridization Capture	Improvement
Sequencing Data Per Sample	340 Gbp	18 Gbp	94% reduction
Total Cost per Sample	$1,740	$205	88% savings
Average Coverage Depth	20× average across genome (~17 Gbp)	25× average on targets (~200 Mbp)	20% deeper on relevant regions

Impact:

- Faster identification of allelic variants associated with key traits.
- Significant cost reduction allows more samples to be analyzed within the same budget.
- High-depth coverage improves confidence in variant calls, enabling trait discovery and precision genomic selection and breeding practices.

Comparative Analysis: Targeted vs. Traditional NGS

Experimental Costs:

Items	Whole Genome Sequencing	Hybridization Capture	Total Savings
Sequencing Cost Per Sample (at $5/Gbp sequencing rate)	$1,740	$90	
DNA Extraction and NGS library prep cost per sample	$40	$40	
Capture kit cost per sample (myBaits Wheat Exome kit, pooling 8 libraries per capture reaction)	n/a	$74	
Total cost for 24 samples	**$41,760**	**$4,896**	**$36,384 (88% reduction)**

Calculations:

Whole genome sequencing: $1,740/sample × 24 samples = $41,760

- Cost per sample = DNA extraction and NGS library prep + Sequencing
- DNA extraction and NGS library prep: $40 per sample
- Sequencing: $1,700 per sample (at $5/Gbp sequencing rate)

Hybridization capture: $204/sample × 24 samples = $4,896

- Cost per sample = DNA extraction and NGS library prep + Capture kit + Sequencing
- DNA extraction and NGS library prep: $40 per sample
- Capture kit: $74 per sample (myBaits Wheat Exome kit, pooling 8 libraries per capture reaction)
- Sequencing: $90 per sample (at $5/Gbp sequencing rate)

Data Volume:

Sequencing amounts	Whole Genome Sequencing (20x read depth)	Hybridization Capture (25x read depth)	Total data reduction
Gbp per sample	340	18	
Gbp per 24 samples	**8160**	**432**	**7,728 Gb (95% data reduction, 20% coverage boost)**

Calculations:

Whole genome sequencing: 17 Gbp per genome × 20× average desired read depth coverage × 24 samples = 8,160 Gbp total

Hybridization capture: 18 Gbp per targeted exome (for 25× average target read depth coverage) × 24 samples = 432 Gbp total

---

Conclusion

Hybridization capture for targeted NGS offers a transformative approach for any NGS project that would benefit from focusing sequencing on key genomic regions of interest, and/or higher coverage of those regions. Large-genome crops like wheat particularly benefit from the advantages of hybridization capture, including:

- Per-sample and overall project costs can be reduced by 88%
- Fewer sequencing runs and more efficient dataset processing can translate to significantly reduced experimental turnaround times
- Coverage depth and variant detection in key genes are enhanced

This strategy empowers breeding programs to accelerate trait discovery and implement precision genomic selection without overwhelming resources.

Recommended Literature

Esposito, S., D'Agostino, N., Taranto, F., Sonnante, G., Sestili, F., Lafiandra, D., De Vita, P. (2022). **Whole-exome sequencing of selected bread wheat recombinant inbred lines as a useful resource for allele mining and bulked segregant analysis**. Frontiers in Genetics, 13. https://doi.org/10.3389/fgene.2022.1058471

Wang, H., Bernardo, A., St. Amand, P., Bai, G., Bowden, R. L., Guttieri, M. J., & Jordan, K. W. (2023). **Skim exome capture genotyping in wheat**. The Plant Genome, 16, e20381. https://doi.org/10.1002/tpg2.20381

Zhang, J., Debernardi, J. M., Burguener, G. F., Choulet, F., Paux, E., O'Connor, L., Enk, J., & Dubcovsky, J. (2023). **A second-generation capture panel for cost-effective sequencing of genome regulatory regions in wheat and relatives**. The Plant Genome, 16, e20296. https://doi.org/10.1002/tpg2.20296