High Throughput Low Pass Whole Genome Sequencing for Genomics Applications in Aquaculture

Sabrina Sauri, Melissa Allen, Dev Mashruwala, Carlos Pulgarin, Jorge Arlacon, Jason Stannard and Klara Verbyla

THE CENTER FOR AQUACULTURE TECHNOLOGIES

What is WGS and Why is it Useful?

Whole Genome Sequencing (WGS) is a method that determines the DNA sequence across an organism's entire genome. Instead of targeting specific genes or markers, WGS captures **genome-wide variation**, including SNPs and other polymorphisms, distributed across all chromosomes.

- Provides unbiased, genome-wide genetic data, avoiding the ascertainment bias of fixed SNP panels
- Captures both common and rare variants, which can contribute to complex traits
- Is highly scalable, especially when implemented as low-pass WGS for large populations

Why Choose Low Pass WGS?

For many aquaculture species, the lack of practical genotyping tools has limited the use of genomic selection in breeding programs. Low-pass whole genome sequencing combined with imputation offers a flexible solution that works across species without the need to create custom marker panels for unique populations.

By enabling programs to evaluate a larger number of animals each generation while improving the return on investment and scalability of the workflow, breeders can estimate genetic merit earlier and with greater confidence. This supports:

- Informed broodstock selection
- Shortens breeding cycles
- Results in steadier genetic progress for traits such as growth, survival, and disease resistance

What is Genomic Selection

Genomic selection is a breeding strategy that uses genome-wide genetic information to predict the breeding value of individuals. Instead of selecting breeders based solely on observed traits or a small number of markers, genomic selection uses thousands to millions of genetic variants spread across the genome to estimate an individual's genetic potential.

Why Genomic Selection is Useful

- Increasing selection accuracy, especially for traits that are difficult, expensive, or time-consuming to measure (e.g., disease resistance, growth rate, feed efficiency)
- Reducing generation interval, because individuals can be selected at an early life stage based on their DNA
- Accelerating genetic gain, leading to faster improvement of economically important traits

Stability of the Technology

As reference genomes are improved, annotations refined, and population-specific reference panels expanded, existing low-pass sequencing data can be reanalyzed and imputed with increasing accuracy—without the need to recollect samples or regenerate libraries.

This allows breeding programs to continuously benefit from advances in genome assemblies, variant databases, and imputation methods over time.

Wet Lab Workflow

Hands-On Workflow Using SeqWell's AgriPrep Kit

1. gDNA is extracted from Cobia tissue using the **Molgen PurePrep** which uses magnetic bead-based chemistries (Molgen PurePrep animal Non-ETOH 0.0 kit) for DNA purifications.
2. The gDNA is normalized across the plate before setting up the reactions
3. 10 min reaction set up time, and thermal cycling program is 78 mins.
4. Individual libraries are pooled together in equal volumes for **a single pool purification** using MAGwise beads

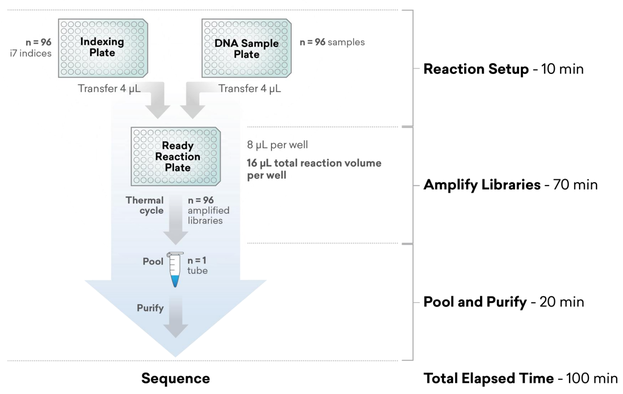
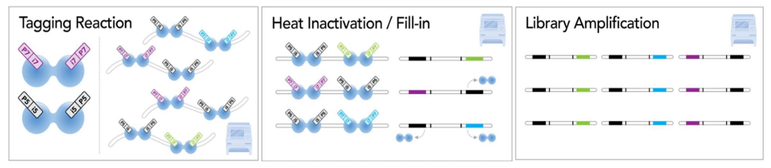
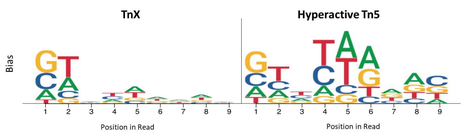


Diagram of Molecular Process



This workflow uses a **SeqWell's TnX transposon system** with a tagmentation strategy to reduce bias during the DNA fragmentation step while producing full-length CDI libraries compatible with the Illumina sequencing platforms such as the **Element AVITI**



MOLGEN seqWell Element Biosciences

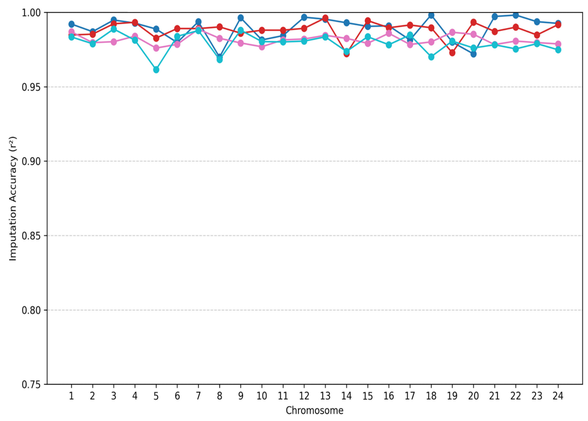
Bioinformatic Workflow

When paired with a high-quality reference, the low-pass data supports **genome-wide genotype recovery**, enabling scalable **genomic selection** in Cobia and providing a framework for other aquaculture species. We constructed a phased reference panel and applied an optimized GLIMPSE2 pipeline to perform variant calling, phasing, and imputation from low-coverage WGS

Across test datasets, the optimized imputation pipeline **achieved ~97% mean genotype concordance** relative to high-density reference data across ~1.3 million SNPs

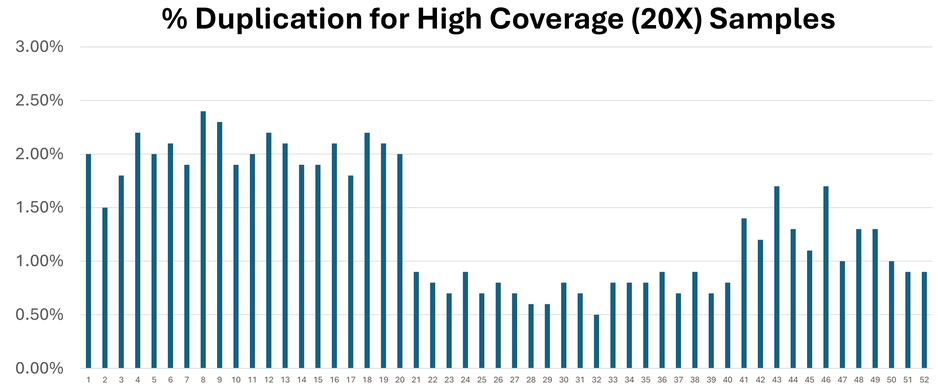
Performance was evaluated across coverage depths (0.2X to 1X) and reference configurations to identify parameters that maximize accuracy and cost-efficiency.

Imputation Accuracy Across Chromosomes in Cobia Samples

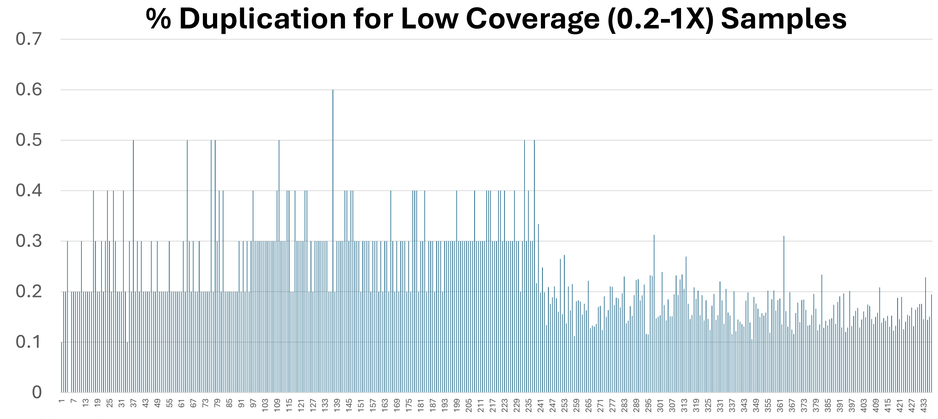


Genome-wide imputation accuracy across all 24 chromosomes for 4 low-coverage whole-genome sequenced cobia samples are all above 95% accuracy

% Duplication for High Coverage (20X) Samples



% Duplication for Low Coverage (0.2-1X) Samples



Duplication rates were low for both high- and low-coverage samples. High-coverage libraries were prepared using traditional enzymatic fragmentation, whereas low-coverage libraries were prepared using SeqWell's TnX tagmentation enzyme