Target Practice: Resolving Methodological Challenges of Baited Targeted Capture of Highly Conserved Genes

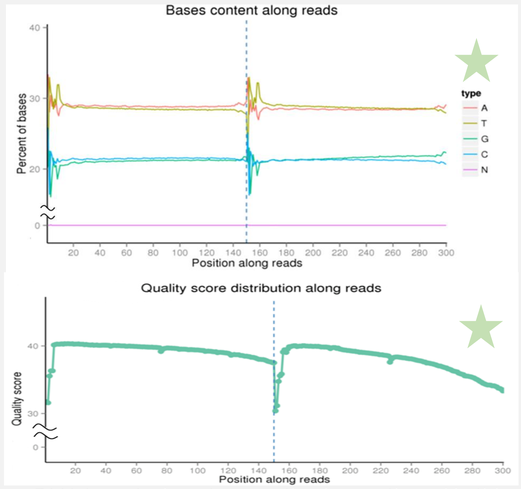
Aguilar MA, Oakley TH

University of California, Santa Barbara

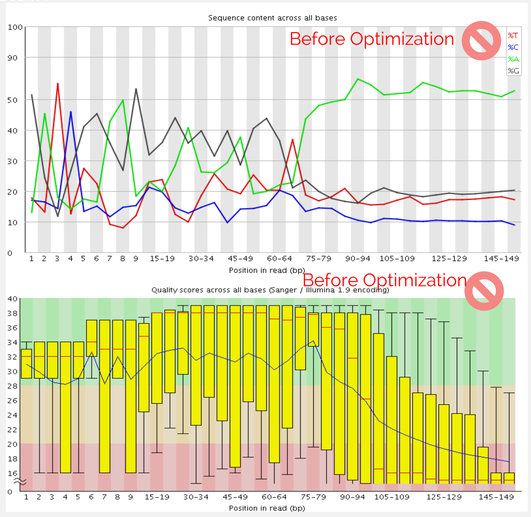
Introduction

While using genomic-DNA targeted gene capture and amplification to build a phylogeny, we encountered challenges within the gene capture procedures. Although gene capture promises an efficient protocol for massively paralleled sequencing, we will introduce several modifications to this technique to give gene capture users more reproducibility and control over this novel yet laborious technique. This poster will follow the work flow of one mollusk tissue sample. The major steps of the gene capture workflow are divided into sections and are characterized by Tapestation DNA traces that represent an optimized sample and a problematic sample. Additionally, the rationale of each new step that circumvents gene capture challenges is described.

Improved Sequencing Results



The gDNA targeted capture methods mentioned in this poster played a vital role in producing high quality sequencing data. The figures above demonstrate that the optimized gene capture library produced substantial data with a high fidelity. Before optimization, the captured genes produced little to no data of low fidelity as seen below.



Objectives and Methods

A — Maximize Yield of DNA to Reduce Excessive Downstream PCR

DNA Phenol/Choloform Extraction and Sonication

B — Detection and Removal of Adapter Dimers and PCR Artifacts

DNA Fragment Size Selection Via Pippen Prep

C — Higher Library Concentration With Fewer PCR Cycles

Illumina Library Preparation via NEBNext Ultra II

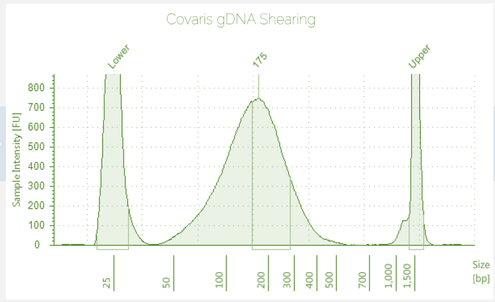
NOTE: We expect the peak from section B to increase in length by 130bp because of the addition of Illumina adapters.

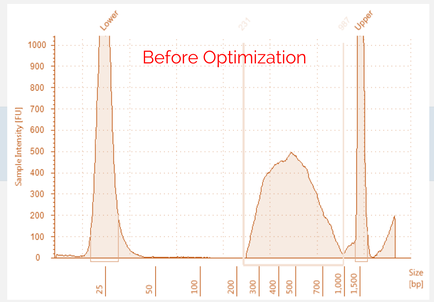
D — Library Amplification Via Replicate PCR in Preparation for Gene Capture

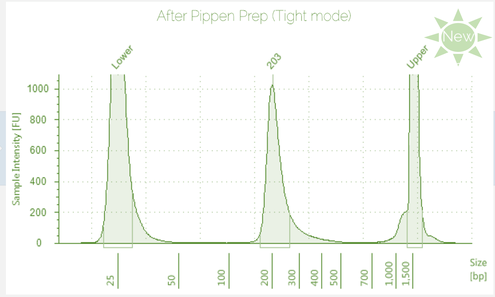
E — Optimize Gene Capture Conditions

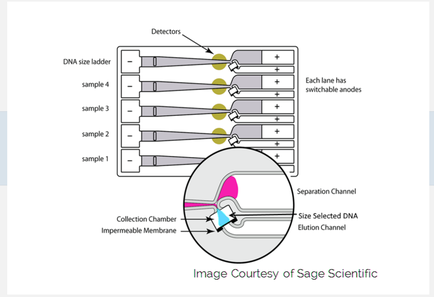
Gene Capture (Slow denaturing and Cool Down to Hybridization Temperature)

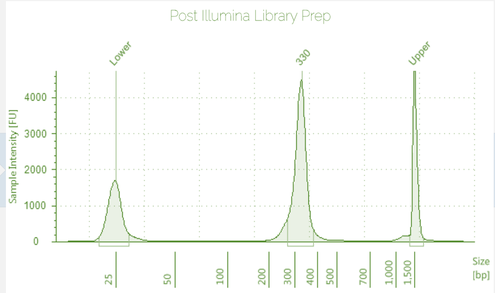
Results (Optimized and Problematic Representations)

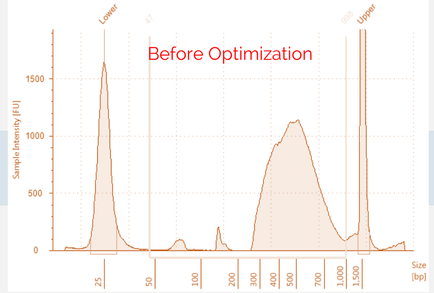


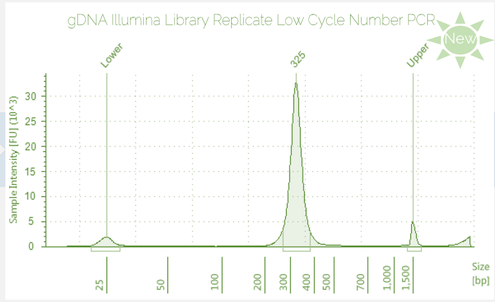


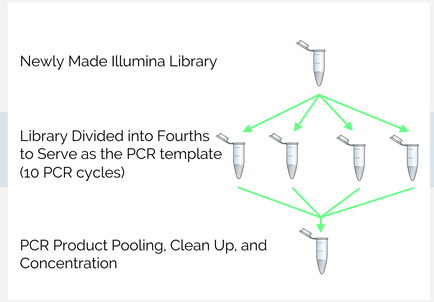


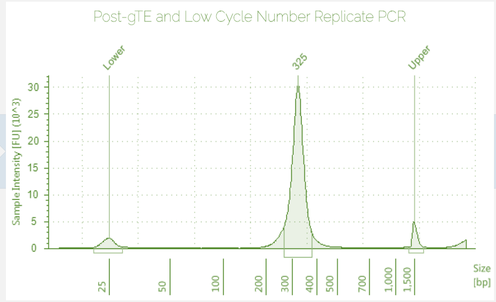


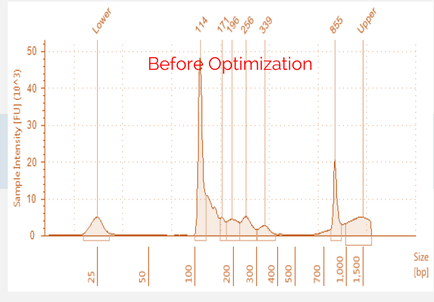












Rationale

Size selection plays a vital role in gene capture because it enables the user to construct a library that is represented by fragments that have approximately 200bp of sequence data.

We want to make ensure that the Illumina library is composed of gDNA fragments that contains an optimal amount of sequence data.

Additionally we want those fragments well represented on the Illumina flowcell when sequencing.

The adapters were diluted 25 fold to minimize adapter-dimer artifacts.

Size selection was skipped entirely within the NEB protocol because it size selects via magnetic beads which has very low specificity in comparison to the Pippen device.

Only 10 cycles of PCR used.

The narrow library peak allows us to easily identify the presence of unexpected high or low molecular weight PCR artifacts out.

It is important to remove unexpected peaks before gene capture to reduce nonspecific capture of PCR artifacts.

The probes are more likely to hybridize with DNA fragments that contain on average 200bp of sequencing data because they predominantly represent the library.

The library denaturing step is vital because it generates a population of ssDNA that the gene capture probes can hybridize to.