

Targeted Sequencing of Genes from Soybean using NimbleGen SeqCap EZ and PacBio SMRT Sequencing

Kevin Eng¹, Anand Sethuraman¹, Denise Raterman², Todd Richmond², Jenny Gu¹, Steve Kujawa¹

¹Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025

²Roche NimbleGen, 500 S Rosa Rd, Madison, WI 53719



ABSTRACT

Full-length gene capture solutions offer opportunities to screen and characterize structural variations and genetic diversity to understand key traits in plants and animals. Through a combined Roche NimbleGen probe capture and SMRT Sequencing strategy, we demonstrate the capability to resolve complex gene structures often observed in plant defense and developmental genes spanning multiple kilobases. The custom panel includes members of the WRKY plant-defense-signaling family, members of the NB-LRR disease-resistance family, and developmental genes important for flowering. The presence of repetitive structures and low-complexity regions makes short-read sequencing of these genes difficult, yet this approach allows researchers to obtain complete sequences for unambiguous resolution of gene models. This strategy has been applied to genomic DNA samples from soybean coupled with barcoding for multiplexing.

EXPERIMENTAL DESIGN

NimbleGen SeqCap EZ Capture Design Panel

Soybean Reference Assembly: *Glycine_max_v2.0* (GCF_00004515.4)
JGI *Glycine max* cultivar Williams 82.

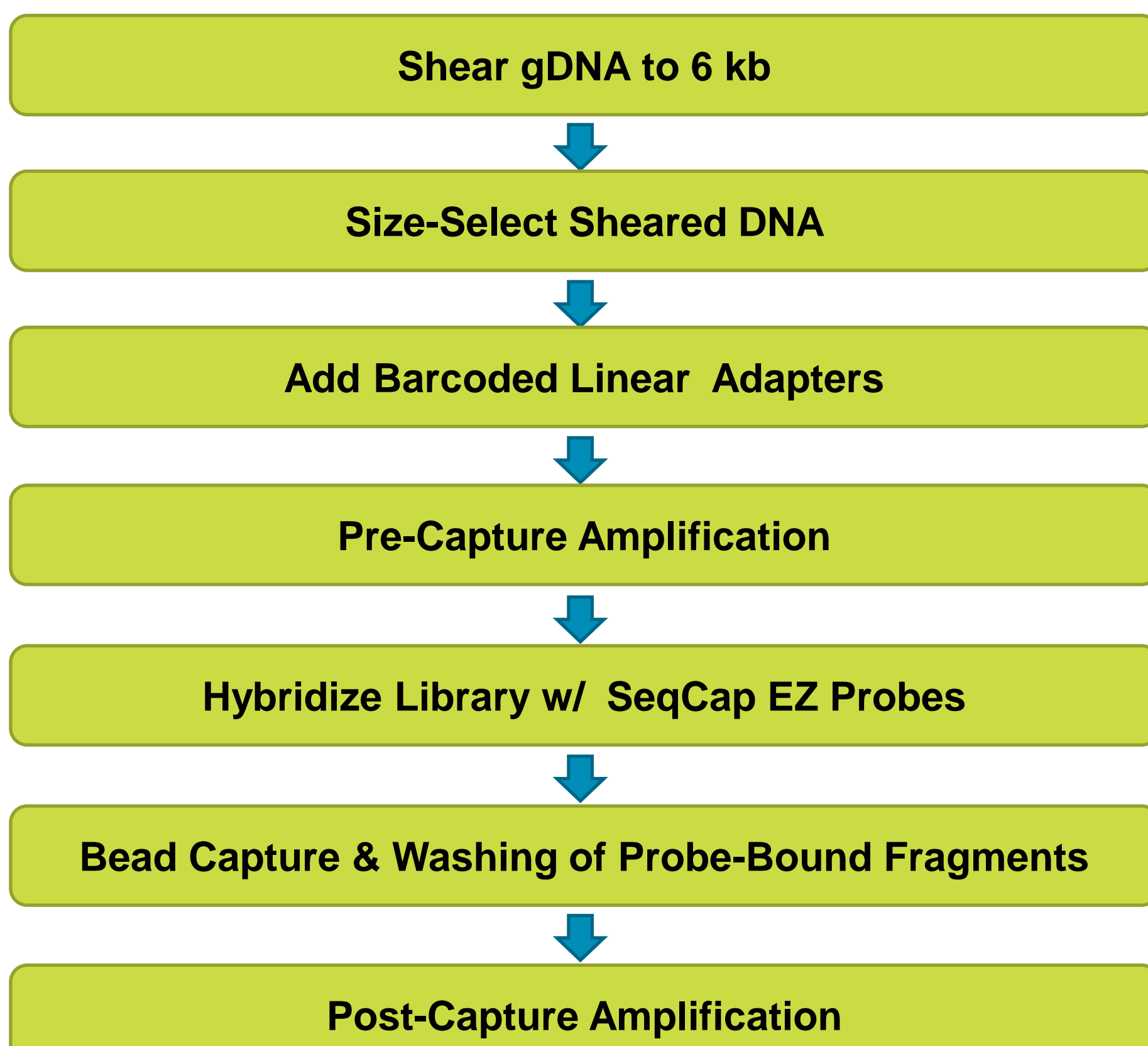
IMMUNOGENES

Accession ID	description	chromosome	start position	end position	gene size	# exons
NC_016088.2	WRKY32 protein	1	7371188	7378764	7576	6
NC_016090.2	NB-LRR type disease resistance protein	3	18612226	18619671	7445	7
NC_016091.2	WRKY45 protein	4	48943779	48947058	3279	4
NC_016091.2	WRKY21	4	48971564	48973995	2431	3
NC_016093.2	TIR-NBS-LRR type disease resistance protein	6	44785526	44803179	17653	10
NC_016094.2	WRKY55 protein	7	5100408	5103289	2881	3
NC_016099.2	TMV resistance protein N-like	12	39885296	39889547	4251	5
NC_016101.2	WRKY49 protein	14	46522674	46526089	3415	5
NC_016102.2	WRKY42 protein	15	14945391	14947423	2032	3
NC_016102.2	CC-NBS-LRR class disease resistance protein	15	15007753	15012607	4854	3
NC_016104.2	WRKY24 protein	17	5778067	5786073	8006	5
NC_016106.2	functional candidate resistance protein KR1	19	9162154	9166787	4633	5
NC_016106.2	NBS-LRR disease resistance protein	19	30576664	30581024	4360	5

DEVELOPMENTAL GENES (FLOWERING)

Accession ID	description	chromosome	start position	end position	gene size	# exons
NC_016089.2	VIN3-like protein 2-like	2	40110289	40115478	5189	4
NC_016095.2	VIN3-like protein 1-like	8	3138161	3148011	9850	7
NC_016100.2	VIN3-like protein 2-like	13	21226990	21232204	5214	5
NC_016103.2	protein HEADING DATE 3A-like	16	31109907	31114981	5074	4
NC_016103.2	brother of FT and TFL1 protein	16	3577864	35779043	1179	5
NC_016104.2	VIN3-like protein 2-like	17	4787557	4792711	5154	5
NC_016105.2	protein TWIN SISTER of FT-like	18	57653696	57659117	5421	4
NC_016105.2	protein HEADING DATE 3A-like	18	57665185	57672971	7786	5

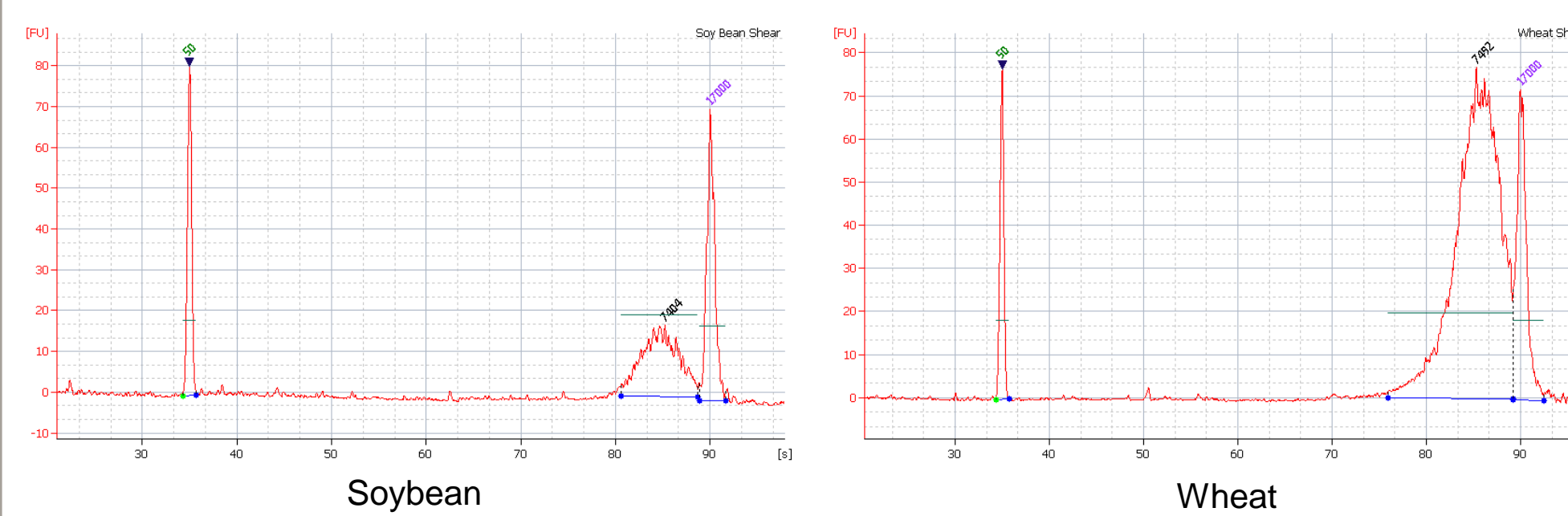
NimbleGen Capture Workflow Overview



NIMBLEGEN CAPTURE RESULTS

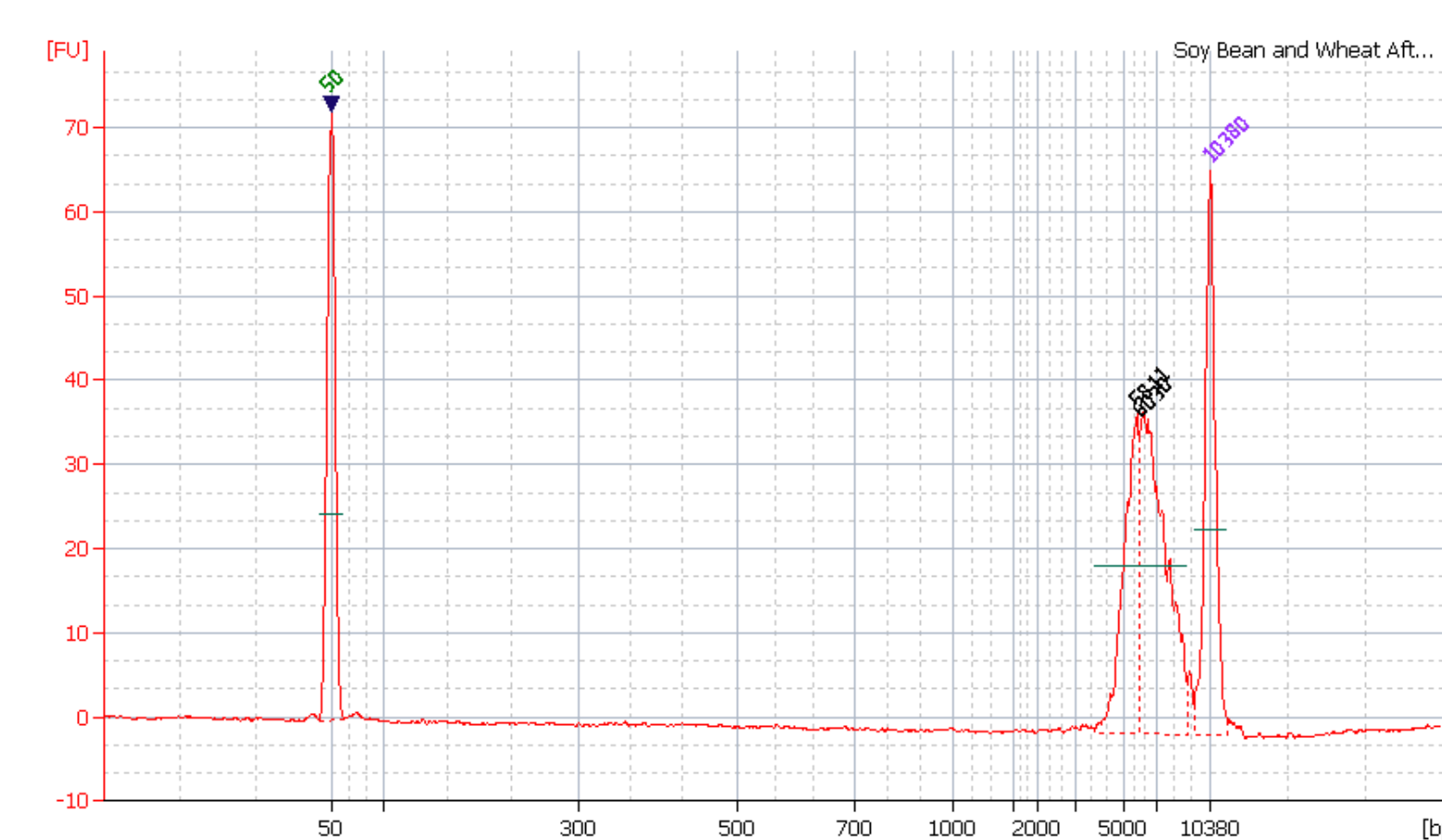
Sheared Genomic DNA

Genomic DNA sourced from Zyagen <https://www.zyagen.com/>



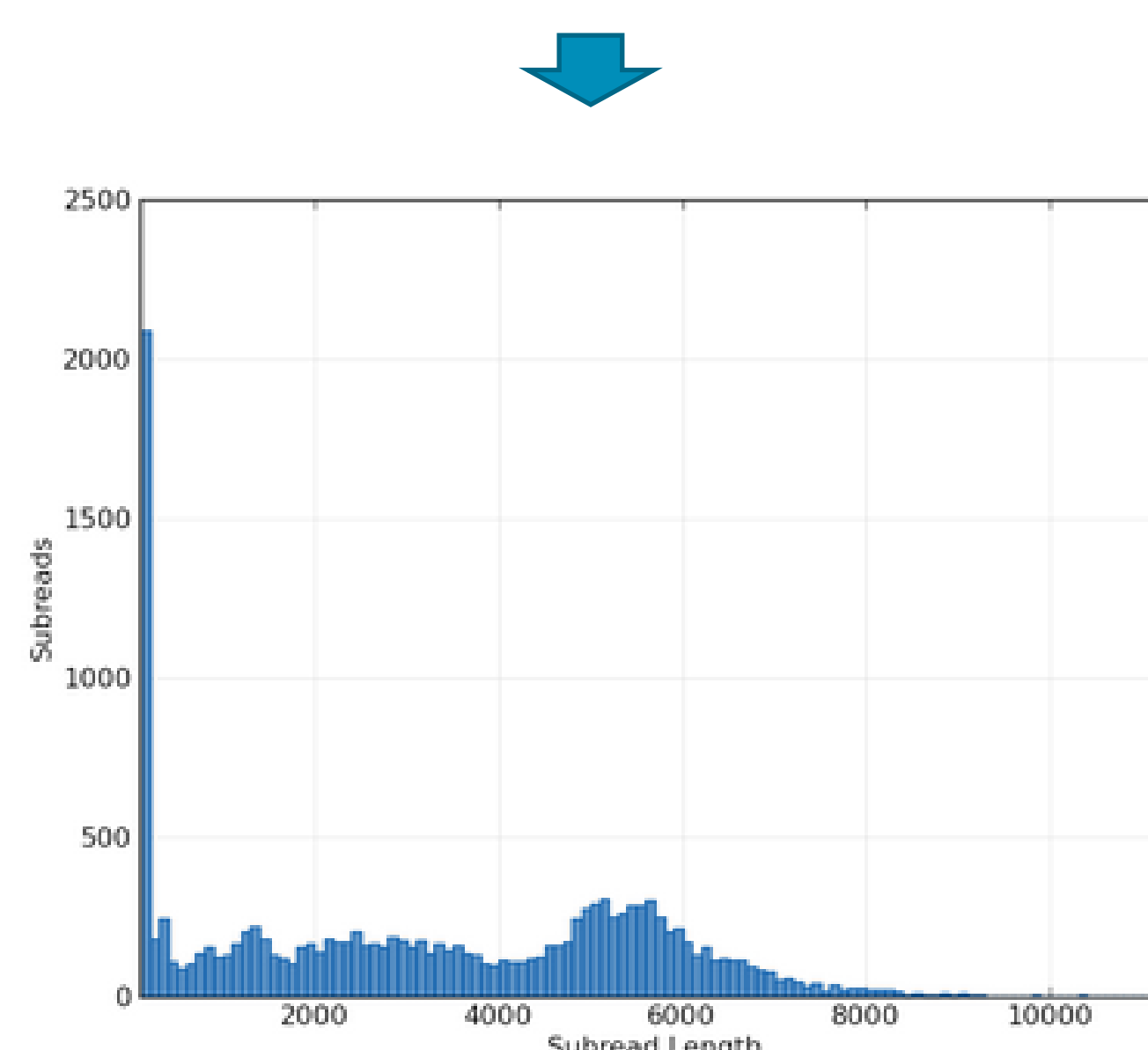
After NimbleGen SeqCap EZ Capture

Pooled Soybean and Wheat genomic DNA with barcodes



After the post-capture amplification step, this sample exhibits a well-defined peak at 6,030 bp. The targeted capture pull-down is now ready for SMRTbell library prep and sequencing.

SMRTbell™ Library Prep & Sequencing (2 SMRT® Cells)

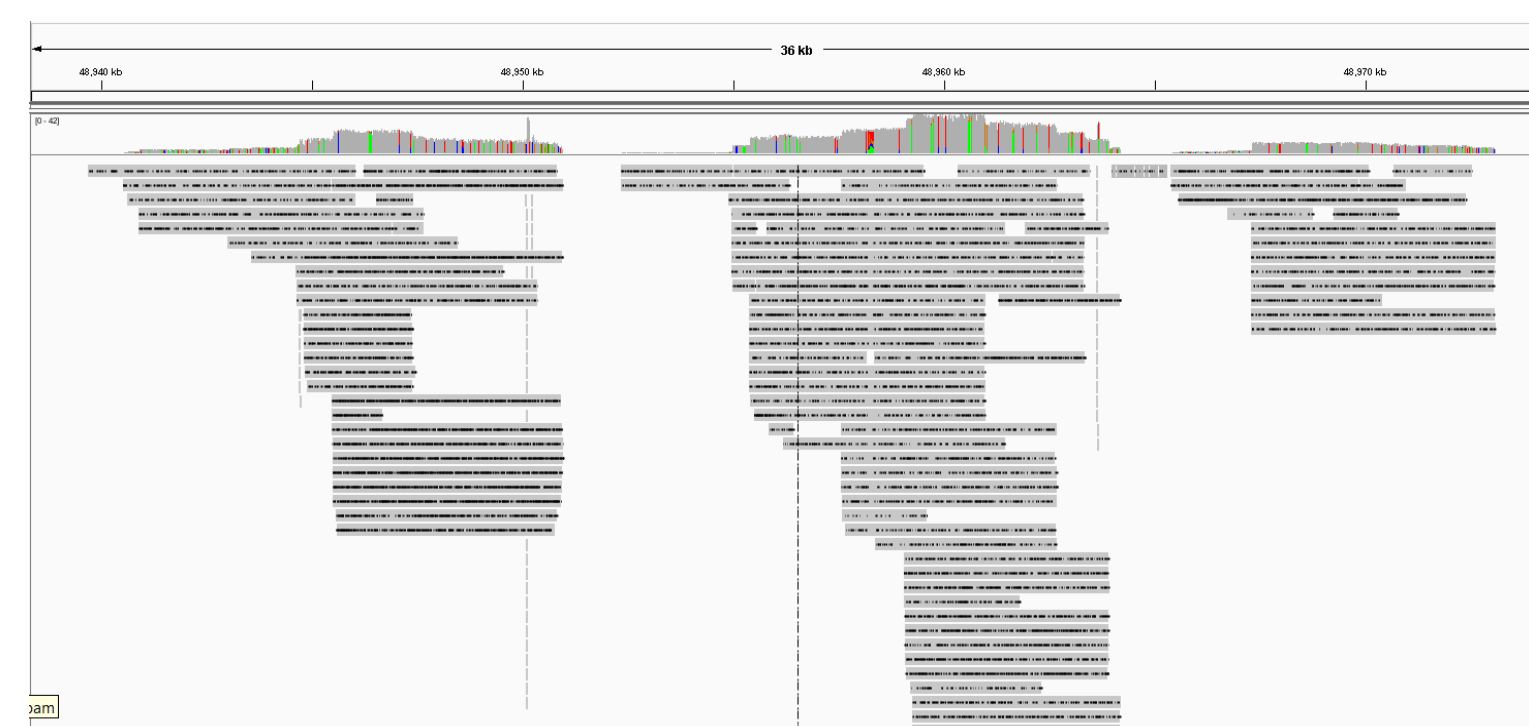


Mapped reads of insert plot from samples shows a primary peak around 5.8 kb and a secondary peak around 3 kb. Loading on the SMRT Sequencing platform was suboptimal and therefore with reduced yield. This will be further optimized.

Generate Reads of Insert / CCS (min 0 full pass, 75% pred. accuracy)

Align to Glycine max v2.0 Assembly

WRKY 45 protein chr4: 48,943,779-48,947,058
WRKY 21 chr4: 48,971,564-48,973,995
Span shown: 36 kb



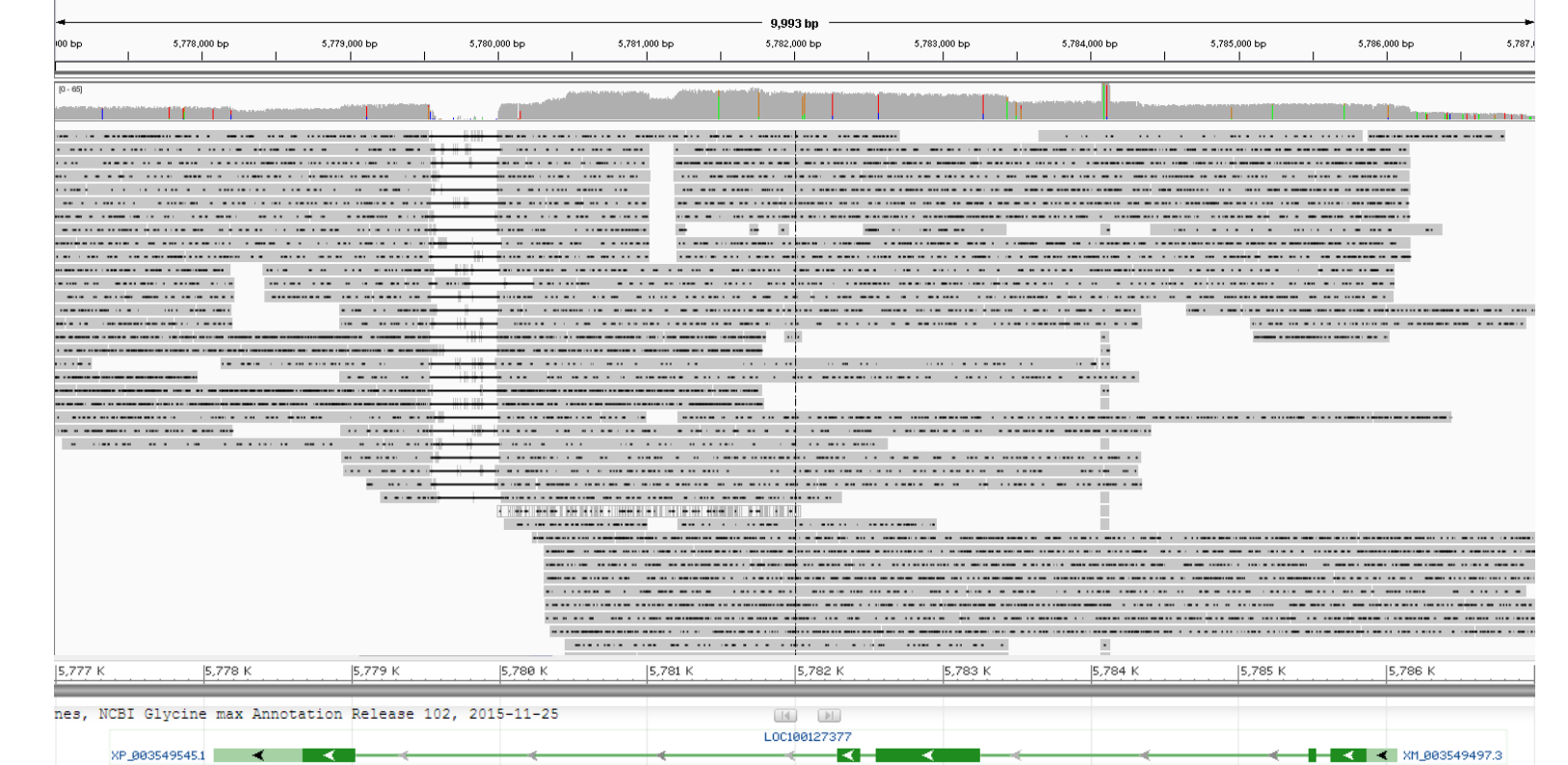
ANALYSIS

IMMUNORESPONSIVE GENES & TRANSCRIPTION FACTORS

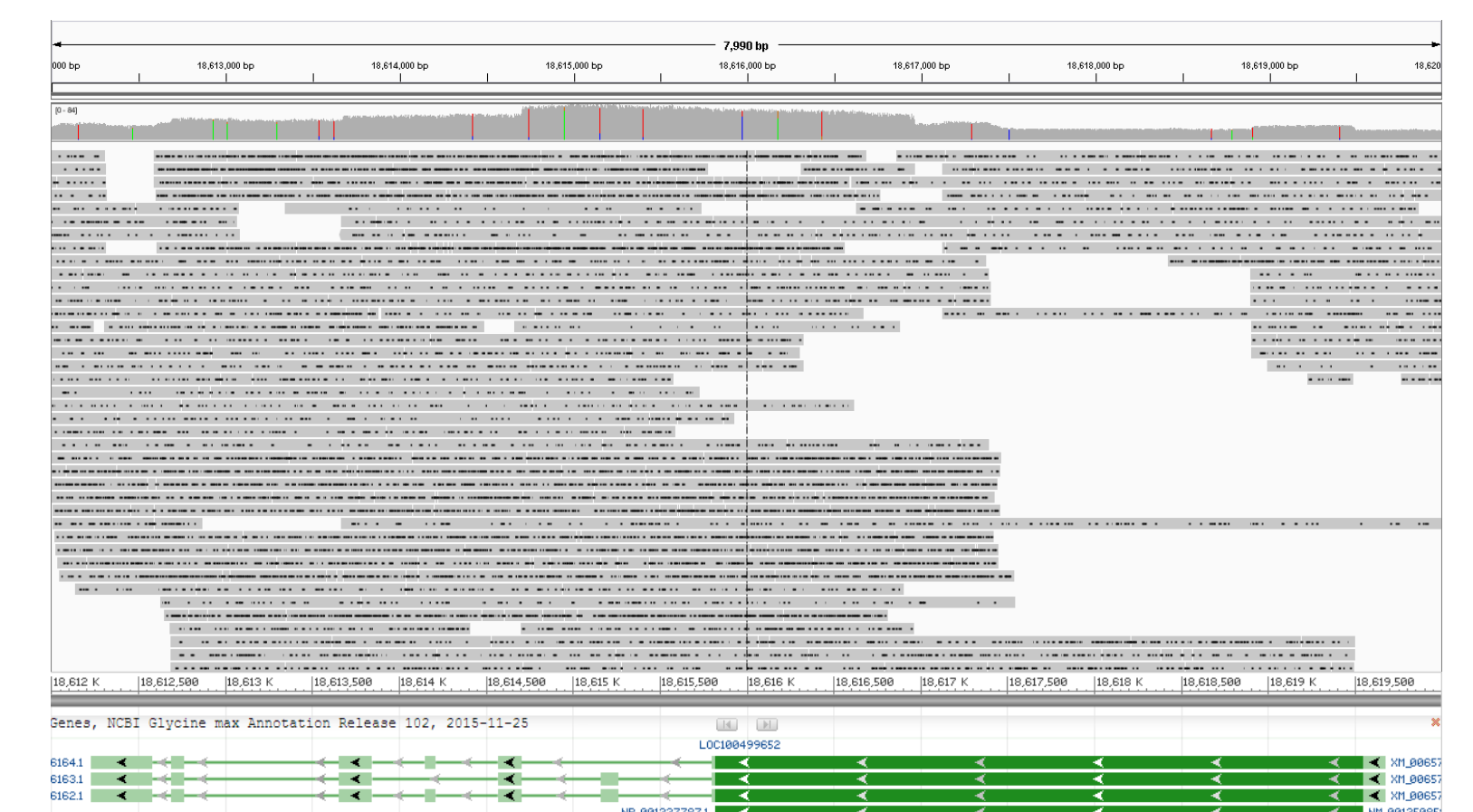
TMV resistance protein N-like chr12: 39,885,296 - 39,889,547
Span shown: 5,000 bp



Probable WRKY transcription factor 32 chr17: 5,778,067 - 5,786,073
Span shown: 9,993 bp



NB-LRR type disease resistance protein chr3: 18,612,226 - 18,619,671
Span shown: 7,990 bp



DEVELOPMENTAL GENES (FLOWERING GENES)

VIN3-like protein 2-like chr2: 40,110,289 - 40,115,478
Span shown: 10 kb



SUMMARY AND RESOURCES

Preliminary data indicated the ability to pull down and sequence >5 kb fragments using NimbleGen SeqCap EZ with good coverage to span entire genes and detect structural variations. Multiplexing of samples for comparative analysis is also possible (data not shown). While success for on-target capture has been demonstrated, this workflow will benefit significantly with further optimization.

PacBio Targeted Sequencing Information Available Here:
<http://www.pacb.com/applications/targeted-sequencing/>