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 Kuwada¹
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_000004515.4)
Glycine max cultivar Williams 82.

IMMUNORESPONSIVE GENES & TRANSCRIPTION FACTORS

TMV resistance protein N-like
chr12: 39,865,286 - 39,889,547
Span shown: 5,000 bp

Probable WRKY transcription factor 32
chr17: 5,778,067 - 5,786,073
Span shown: 9,990 bp

DEVELOPMENTAL (FLOWERING GENES)

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

ANALYSIS

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

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

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

Align to Glycine max v2.0 Assembly

NRK45 protein
chr4: 48,943,779-48,947,058
NRK21
chr4: 48,971,564-48,973,995
Span shown: 36 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/