PO0717

SMRT- RenSeq for NLR resistance gene characterization in *Arachis*, *Glycine* and *Musa* species

Gabriel Sergio Costa Alves[1], Ana Paula Zotta Mota[2], Fernando Campos de Assis Fonseca[1], Amanda Cristina de Araújo[1], Michelle Guitton Cotta[1], Roberto Coiti Togawa[2], Marcos Mota Do Carmo Costa[2], Orzenil Bonfim da Silva Jr[2], Fabrício Arraes[2], Maria Fatima Grossi de Sá[2], Ana Cristina Miranda Brasileiro[2], Jacob Enk[3], Tianying Lan[3], Patrícia Messenberg Guimaraes[2], Robert Neil Gerard Miller[1]

[1]Universidade de Brasília, Brasília, Brazil, [2]Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil, [3]Arbor Biosciences, Ann Arbor, MI, USA

INTRODUCTION

The most abundant plant disease resistance (*R)* genes comprise a family that encode intracellular multidomain receptor proteins with a stereotypical nucleotide binding site (NBS) and leucine-rich repeat (LRR) domain (NLRs) (Figure 1). These intracellular receptors recognize directly or indirectly pathogen effector proteins, initiating effector-triggered immunity (ETI). Conserved motifs in such receptors across diverse plant taxa offer a means for their accelerated isolation across target or uncharacterized plant species.

Here, R-gene enrichment sequencing (RenSeq) was employed for reduced complexity selective sequencing of NLR gene sequences in resistant accessions of the species *Arachis stenosperma*, *Glycine max* and *Musa acuminata*.

MATERIALS AND METHODS

TARGET IDENTIFICATION AND PROBE DESIGN

Conserved motifs in domains in NLR genes from reference genomes *M. acuminata ssp. malaccensis* (DH-Pahang V2), *A. duranensis* (Aradu 1.0) and *G. max* (Williams 82) were used to build species-specific HMM profiles for screening of genomic and transcriptomic sequence data for pfam NLR signature domains (TIR, CC, RPW8, NB-ARC and LRR) from the target disease resistant genotypes *M. acuminata ssp. burmannicoides* var. Calcutta 4, *G. max* (PI595099) and in the wild species *A. stenosperma* (Figure 1).

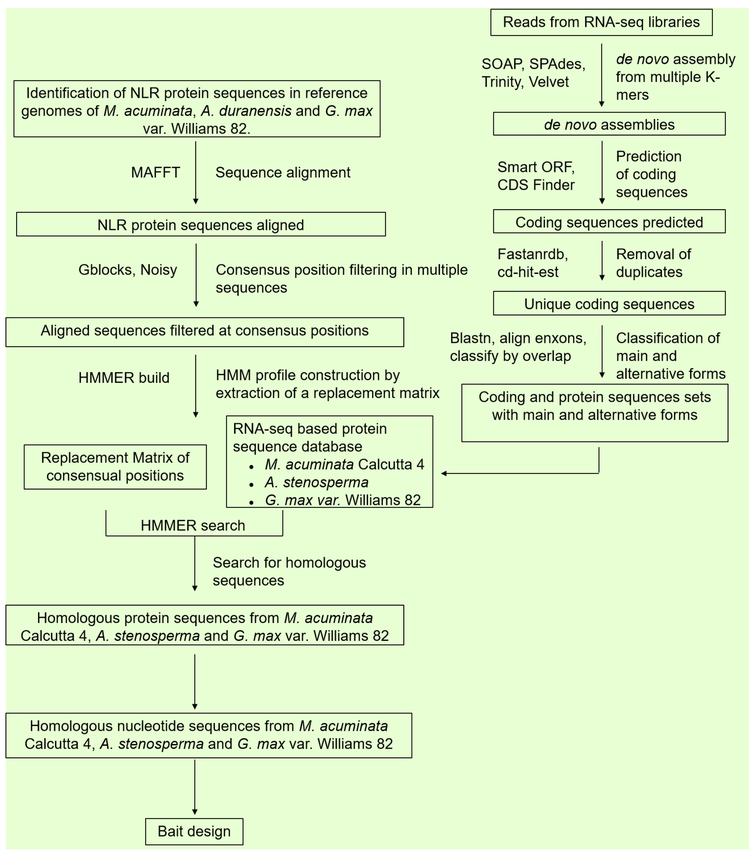

Figure 1: Flowchart summary of bioinformatic pipeline employed for NLR signature domain identification for specific RNA bait design

TARGET ENRICHMENT WITH SMRT RenSeq

Using biotinylated RNA baits designed for target enrichment from gDNA (myBaits®, Arbor Biosciences, USA), a combined approach of R-gene sequence enrichment and single-molecule real-time sequencing (SMRT RenSeq) was employed for accurate *de novo* assembly of NLR gene repertoires (Figure 2). Enriched libraries were prepared for long read PacBio sequencing, with size selection conducted using the Blue Pippin System (Sage Science, MA, USA) for 3 kb and longer fragments. Samples were multiplexed and PacBio Sequel SMRT sequenced using v. 3 chemistry.

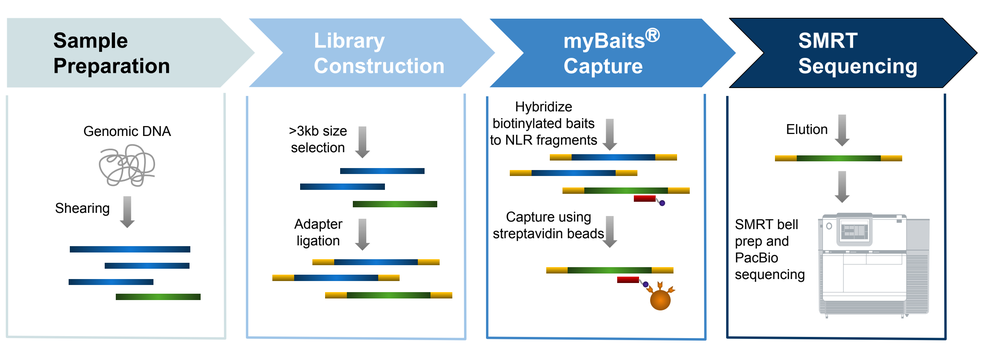

Figure 2: Schematic illustration of the SMRT RenSeq workflow

DATA ANALYSIS AND NLR RECONSTRUCTION

Braker was employed for automated prediction of protein coding gene structures (Hoff et al., 2019), with NLR gene sequences further verified using NLR annotator for predicted stereotypical domains and motifs (Figure 3) (Steuernagel et al., 2018).

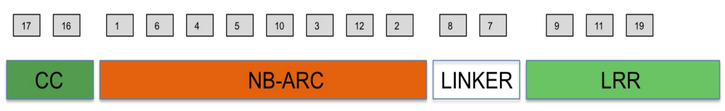

Figure 3: Scheme depicting known conserved amino acid motifs (numbered) within plant NLR immune receptor domains, adapted from Steuernagel et al., 2018

RESULTS

SMRT RenSeq enabled detection of full-length NLR loci for *A. stenosperma*, *G. max* PI595099, and *M. acuminata* Calcutta 4, with 450, 408 and 114 confirmed as complete genes, respectively, for each species.

Phylogenetic analysis was employed to order NLR genes in terms of sequence relationship and to group into sub-classes, colour-coded on the Phylogenetic tree (Figure 3) and summarized in Table 1.

	CC	TIR	xNL	Total
Arachis stenosperma	246	117	87	450
Glycine max	155	156	97	408
Musa acuminata	101	0	13	114

Table 1: Summary of complete NLR genes and sub-class members identified for *Arachis stenosperma*, *Glycine max* and *Musa acuminata* Calcutta 4, based on phylogenetic analysis.

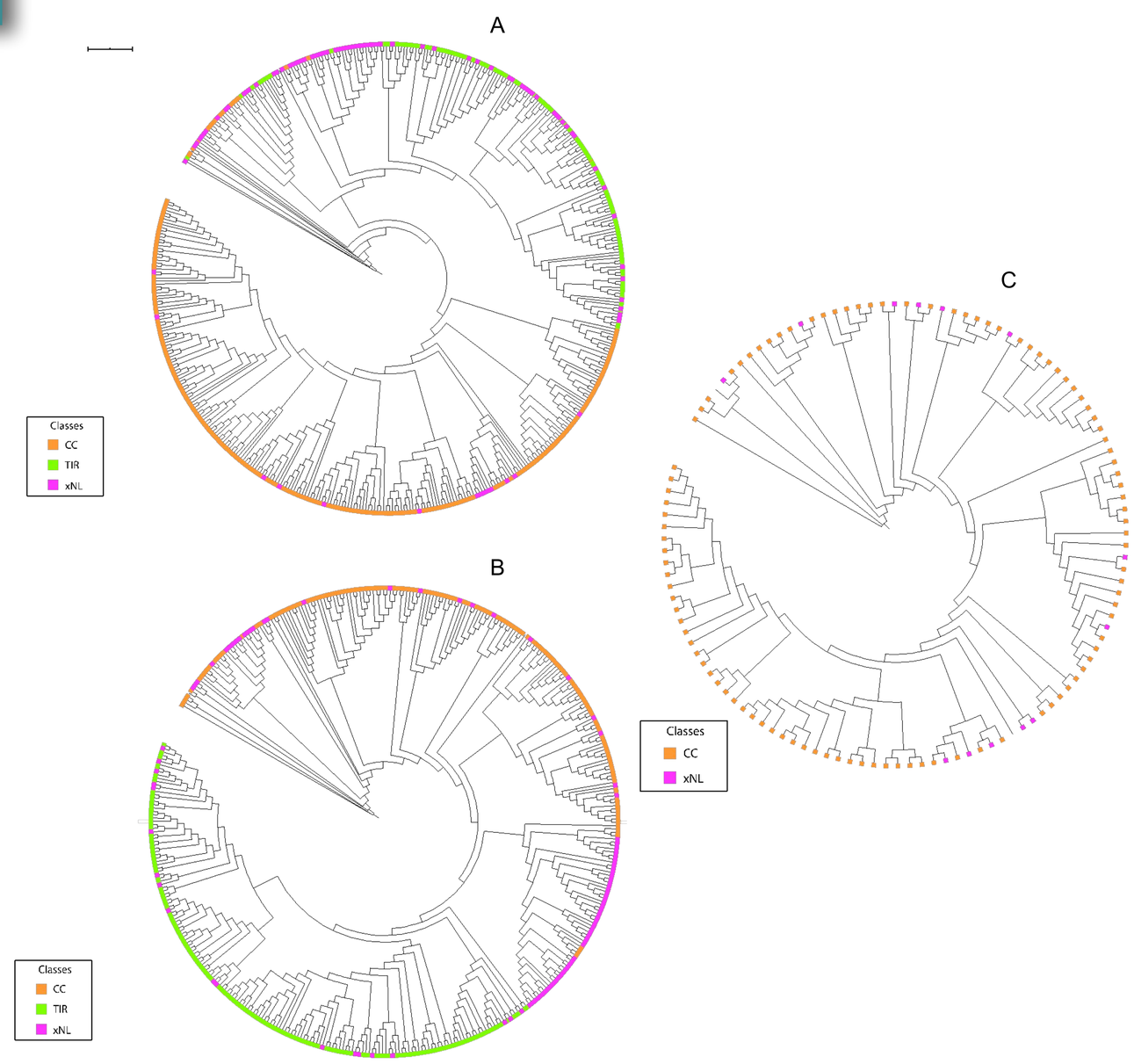

Figure 3: Phylogenetic analyses of the NB_ARC domains from predicted NLR genes for *Arachis stenosperma (A)*, *Glycine max* PI595099 (B) and *Musa acuminata* Calcutta 4 (C). Evolutionary relationships were inferred using the Maximum likelihood method. Analysis involved 450 amino acid sequences from *Arachis stenosperma*, 408 from *Glycine max* PI595099 and 114 from *Musa acuminata* Calcutta 4. Evolutionary distances were calculated using the JTT method. Phylogenetic analysis was conducted using the program RAxML.

CONCLUSIONS

This large-scale enrichment sequencing approach enables accurate and exhaustive characterization of NLR genes in resistant accessions for the genera *Arachis*, *Glycine* and *Musa,* of relevance for global food security.

Full length NLR genes are a resource for functional gene cloning and editing for the engineering of durable disease resistance.

References

Hoff, K.J., Lomsadze, A., Borodovsky, M. and Stanke, M. 2019. Whole-Genome Annotation with BRAKER. Methods Mol Biol. 1962:65-95, doi: 10.1007/978-1-4939-9173-0_5.

Steuernagel et al., 2018. Physical and transcriptional organisation of the bread wheat intracellular immune receptor repertoire. bioRxiv preprint. http://dx.doi.org/10.1101/339424.