NLR-ANNOTATOR: A TOOL FOR DE NOVO ANNOTATION OF INTRACELLULAR IMMUNE RECEPTOR REPertoire

Nucleotide-binding domain Leu-rich repeat (NLR) proteins serve as intracellular immune receptors in plants to recognize different types of effectors delivered by pathogens from all kingdoms. NLR genes form the largest plant-disease–resistance gene family. NLR proteins share conserved NB-ARC domains at the N terminus and variable Leu-rich repeat domains at the C terminus for effector targeting specificity (Monteiro and Nishimura, 2018). During the arms race with pathogens, plant genomes have accumulated numerous NLR genes with great diversity, including copy-number variations across species, structural variation, and single-nucleotide polymorphisms, especially at effector targeted C-terminal domains. The NLR gene repertoires of plants represent valuable agronomic traits for durable and broad-spectrum resistance in breeding programmes of plants represent valuable agronomic traits for domestication. Another difficulty is that automated gene annotation tools rely on transcriptomic sequences upstream and downstream for additional NLR-associated motifs, such as coiled-coil domains or Leu-rich repeat domains. By combining all reported NLR loci, NLR-Annotator can thereby annotate the NLR repertoire for a whole genome (https://github.com/steuernb/NLR-Annotator; Fig. 1).

The quality of any annotation is evaluated by two parameters: sensitivity and specificity. The sensitivity of an annotation in this case is determined as the ratio of identified NLR genes to all NLR genes. The specificity of the annotation is the ratio of correctly identified NLRs to all identified NLRs. The NLR-Annotator exhibits both high sensitivity and specificity when applied to the Arabidopsis (Arabidopsis thaliana) genome, which is usually used as a gold standard because of its well-characterized NLR genes (Meyers et al., 2003; van de Weyer et al., 2019). Comparative analysis of NLR repertoires across crop cultivars and other plant species requires the universal application of NLR gene annotation tools. Steuernagel et al. (2020) successfully applied NLR-Annotator to eight economically important crop genomes, including one food and industrial resource crop, soybean (Glycine max); two cereal-related crops, maize (Zea mays) and purple false brome (Brachypodium distachyon); and five horticultural crops, including cucumber (Cucumis sativus) and potato (Solanum tuberosum). These demonstrate the broad applicability of NLR-Annotator across diverse plant taxa in phylogenetic construction using NLR genes identified with the same standard.

With NLR-Annotator in hand, the group next explored the NLR repertoire in T. aestivum by annotating NLR genes in a Chinese Spring cultivar with a high-quality genome assembly (International Wheat Genome Sequencing Consortium, 2018). They identified 3,400 NLR loci and 1,560 complete NLRs, and they found some intriguing features of the NLR repertoire in the bread wheat. For example, NLR loci proved to distribute predominantly across all chromosomes at their telomere regions, and half of them cluster together. The genomic arrangement pattern likely links with the evolutionary mechanisms underlying NLR gene expansion within a species. There are ~8% of proteins across the whole genome with integrated domains that encode proteins acting as a
decoy or bait. The decoy proteins specialize solely in perception of the effector by the NLR protein by mimicking the effector targets and they are novel candidate genes for crop resistance breeding (van der Hoorn and Kamoun, 2008). The analysis in wheat also revealed the sequences of the NLR genes, with their functional and evolutionary relationships. Finally, the wheat NLR repertoire allowed the exploration of whole-genome expression profiles of NLR genes, which showed a majority of NLRs with low-expression and stress-induction features.

In brief, with the NLR-Annotator pipeline of Steuernagel et al. (2020), researchers have a tool for the rigorous and reproducible annotation of NLRs across plant taxonomic clades. The species-level identification and description of NLR repertoires enable the phylogenetic construction of NLR gene families, which will provide insights on functional and evolutionary relationships among NLR genes. This knowledge will be essential for future advances that harness NLRs in economically important crops through breeding and genome editing.
LITERATURE CITED


