Plants have two kinds of pathogen recognition receptors: extracellular receptor-like kinases and proteins (RLKs and RLPs) and intracellular nucleotide-binding leucine-rich repeat (NLR) receptors. NLRs comprise three main domains: a central nucleotide binding domain (NB-ARC) that mediates receptor oligomerization upon activation, a C-terminal LRR domain that defines receptor specificity, and an N-terminal domain that mediates immunity. Based on the latter domain, the NLRs are subdivided into three monophyletic groups: RNLs (Resistance to Powdery Mildew), CNLs (Coiled-Coils), and TNLs (Toll/Interleukin-1 Receptor homology). NLRs can be sensors or signal transducers. As sensors, NLRs can recognize pathogens by directly binding the effectors, by recognizing the effector’s action on other proteins, or by recognition of modifications to a non-canonical NLR domain. Continuous generation of NLR diversity is required to keep up with a range of rapidly evolving pathogens.

Shannon entropy analyses of sequence alignments have been used successfully to identify highly variable residues that determine immunoglobulin receptor specificity (Shenkin et al., 1991). In this study, Prigozhin and Krasileva (2021) used this existing theory to investigate sequence diversity in pan-NLRomes. The authors described a new methodology to study pan-NLRomes using existing data for both Arabidopsis thaliana accessions (Van de Weyer et al., 2019) and Brachypodium distachyon lines (Gordon et al., 2017) that complemented the previous molecular evolution analyses.

First, the NLRs were newly classified based on their shared NB-ARC domains using phylogeny-based approaches. By this, the authors circumvented the problem of losing robust allelic series for each gene that can occur with distance-based clustering techniques. The resulting subclades differed in size and sequence diversity demonstrating that closely related NLR sequences evolve at different evolutionary rates throughout the NLR family.

Next, by using Shannon entropy, the authors analyzed the diversity in subfamily alignments and defined the highly variable NLRs (hvNLRs). In these receptors, the highly variable residues are expected to form the target binding sites. In plants, LRR domains are known to carry recognition specificities, and indeed, highly variable residues were mainly located there, although, their exact location varied between NLR subfamilies. An easy-to-read matrix representation of Shannon entropy scores mapped on to the LRR repeats (rows: repeat units and columns: variable positions in the canonical LXXLXLC repeat) allowed skipping structure prediction as shown for RPP13 (see Figure 1).

To validate the binding site predictions, the authors focused on RPP13, both well-known NLR disease resistance genes. Using synthetic RPP13 variants that transferred highly variable residues (entropy > 1.5) between Col-0 and Nd alleles, and the ATR13 effector as a target in transient tobacco assays, the authors confirmed the requirement for the high entropy residues for target recognition.

Finally, a similar analysis on Brachypodium lines demonstrated that phylogeny-based NLR analysis combined with Shannon entropy matrices can also be applied to non-model plants. Future automated algorithms will further ease this computational pan-NLRome analysis tool. Complementary structural and mutational experiments to confirm the entropy-based target binding sites predictions shall help convince other researchers to use this methodology.
**Figure** RPP13's LRR domain 2-D matrix representation and homology model. (Left) Shannon entropy scores and amino acid residues of hvNLRs RPP13 mapped onto the 2-D surface representation including additional five amino acids on either side of the core repeat. (Right) RPP13 LRR domain homology model colored by Shannon entropy scores: low (light blue) to high entropy (dark blue). Adapted from Prigozhin and Krasileva (2021), Figures 3, 5.

**References**


