

# Variation patterns

DOI:

10.1038/nrg2198

A thorough analysis of genetic variation in natural populations is crucial to identify the variants that affect phenotypes, and to determine the evolutionary forces underlying such diversity. Clark, Weigel and colleagues now provide a comprehensive polymorphism resource for *Arabidopsis thaliana* that paves the way for both functional and evolutionary studies. Putting this resource to use, Nordborg and colleagues have explored the pattern of linkage disequilibrium (LD) in this model organism, setting the stage for future genome-wide association studies.

In the first study, the authors used a high-density array resequencing approach to survey sequence variation in 20 wild accessions (strains) of *A. thaliana*. They detected more than a million non-redundant SNPs with various levels of confidence using two complementary methods (a model-based algorithm and a machine-learning method). A set of 648,570 SNPs that were detected at a false discovery rate of about 2% was then selected for further analysis. As well as SNPs, they also identified long tracts of polymorphic sequences that corresponded to clusters of SNPs and deletions.

To examine possible functional consequences of the variation that they uncovered, the authors focused on the 26,541 annotated protein-coding

genes in *A. thaliana*. They found that 1,614 of them contain at least one SNP with a large effect on the gene-product integrity (for example, SNPs that insert or remove stop codons or that affect splicing sites), and that 1,191 genes include deletions or highly polymorphic sequences.

Interestingly, changes with large effect occurred with a higher frequency in several gene families. These included nucleotide-binding leucine-rich repeat (NB-LRR) genes and receptor-like kinase (RLK) genes, which are both implicated in strain-specific resistance to pathogens. For these gene families, an analysis of allele frequencies for different types of polymorphism showed that both non-synonymous and synonymous variants are skewed towards high frequency compared with what is observed for other gene families. The unusual variation in NB-LRR and RLK genes might reflect the action of balancing selection, possibly as a result of regional adaptation to the biotic environment.

Zooming out to the chromosomal level revealed that variation has a non-random distribution along the chromosomes. For some regions, including those with high NB-LRR gene number, the authors observed high polymorphism levels, which they suggest result from multiple factors, including linkage to

polymorphisms that are maintained by balancing selection. By contrast, for several regions polymorphism was very low, and long tracts of extended haplotype sharing indicated the action of recent positive selection.

In the second study, Nordborg and colleagues analysed the LD pattern in 19 accessions of *A. thaliana*. They found that, similar to humans, LD decreases within 10 kb, probably owing to a balance between the large size of the worldwide population and its high rate of selfing. LD extent varies widely across the genome, with a population recombination rate that correlates with the level of polymorphism and GC content. The authors show that a genotyping array designed on the basis of these results should have more than adequate coverage for genome-wide association mapping.

What has shaped LD in *A. thaliana*? The authors suggest that both recombination and selection have played a part. They found evidence for recombination hotspots, which were preferentially located outside gene-coding regions, and found that LD is more extensive when flanking non-synonymous SNPs than when flanking synonymous SNPs, which indicates the effect of underlying directional selection.

Altogether, these findings represent an important step towards a better understanding of the forces that shape variation in *A. thaliana* as well as the functional consequences of such diversity.

Francesca Pentimalli



DIGITAL VISION

**ORIGINAL RESEARCH PAPERS** Clark, R. M. et al. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338–342 (2007) | Kim, S. et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet.* 5 August 2007 (doi:10.1038/ng2115)

**FURTHER READING** Gaut, B. S. et al. Recombination: an underappreciated factor in the evolution of plant genomes. *Nature Rev. Genet.* **8**, 77–84 (2007)