IDENTIFICATION OF GENOMIC REGIONS CARRYING A CAUSAL MUTATION IN UNORDERED GENOMES

Pilar Corredor Moreno

The Sainsbury Laboratory, Norwich Research Park, Norwich, UK, NR4 7UH

ABSTRACT

Whole genome sequencing using high-throughput sequencing (HTS) technologies offers powerful opportunities to study genetic variation. Mapping the mutations responsible for different phenotypes is generally an involved and time-consuming process so researchers have developed user-friendly tools for mapping-by-sequencing, yet they are not applicable to organisms with non-sequenced genomes. We introduce SDM (SNP Distribution Method), a reference independent method for rapid discovery of mutagen-induced mutations in typical forward genetics screens. SDM aims to order a disordered collection of HTS reads or contigs such that the fragment carrying the causative mutations can be identified. SDM uses typical distributions of homozygous SNPs that are linked to a phenotype-altering SNP in a non-recombinant region as a model to order the fragments. To implement and test SDM, we created model genomes with SNP density based on *Arabidopsis thaliana* chromosome 1 and analysed fragments with size distribution similar to reads or contigs assembled from HTS sequencing experiments. SDM groups the contigs by their normalised SNP density and arranges them to maximise the fit to the expected SNP distribution. We analysed the procedure in existing data sets by examining SNP distribution in recent outcross [17, 18] and back-cross experiments [19, 20] in *Arabidopsis thaliana* backgrounds. In all the examples we analysed, homozygous SNPs were normally distributed around the causal mutation. We used the real SNP densities obtained from these experiments to prove the efficiency and accuracy of SDM. The algorithm could successfully identify small sized (10-100 kb) genomic regions containing the causative mutation.