Identification of Natural Selection in Mayetiola destructor SNP Markers from Sequencing Data

Jung, Luann

This research presents a computational pipeline for single nucleotide polymorphism (SNP) analysis. Due to the scientific potential of the complex and multifaceted information provided by SNP data, it is being generated at an unprecedented speed. However, traditional analysis of SNPs is lacking in both efficiency and conclusiveness. This study creates a computational tool consisting of a linear series of steps - a pipeline - that streamlines the processes of both SNP analysis and gene ontology retrieval. Using principal component analysis, Mahalanobis test statistics, False Discovery Rate control, and functional network creation, the pipeline takes thousands of SNPs as input for analysis and then reports structured information for visualization of relationships and generation of targets for further study. Consequently, adaptations and natural selection can be connected to specific genes. An application of the pipeline (as demonstrated in this study), is the analysis of 7039 Mayetiola destructor (Hessian fly) SNPs to identify gene pathways which lead to differences in fly virulence. The data encompasses three different biotypes and 288 total flies. The output of the pipeline includes the identified significant SNP markers, the gene surrounding each identified SNP, any functional pathways of the genes, functional networks of gene-pathway relationships, and insight into how certain functional pathways are related to differences in phenotype. This tool makes gene identification using SNP data an efficient, automatic process and helps to pinpoint targets for further experimental investigation. Wide application of this tool can drastically accelerate discovery of novel genes using SNP data.

Awards Won:

Monsanto Company: Third Award of $1000