

6th International Conference on

Bioinformatics & Systems Biology

August 22-23, 2016 Philadelphia, USA

Ensembl plants - Tools for visualising mining and analysing crop genomics data

Dan Bolser

European Bioinformatics Institute, UK

Ensembl plants are a genome centric platform for visualisation and analysis of plant genomics data. It hosts assembly, sequence, expression, variation and comparative datasets for a growing number of plant species (currently 39) covering a range of economically important crops, including brassica, tomato, grape, barley, potato, maize and wheat, and taxonomically diverse model organisms. The web-based genome browser visually integrates sequence and assembly information with genes, markers, probes, repeats and other public or user-supplied datasets. It includes a web-based data mining tool, allowing specific sets of data to be queried and downloaded for offline analysis. In addition to the browser, all data can be accessed computationally via extensive Perl and REST APIs and is available for FTP download or direct database access.

dbolser@ebi.ac.uk

Estimation of missing values for gene interaction data coming from high throughput technologies

Gashaw Mekuria

University of Tampere, Finland

Advancements in high-throughput genetic screening technologies have enabled us to systematically study how gene interactions between pairs of genes can affect phenotypes of certain traits. However, these advancements also pose other challenges to researchers in the management and analysis of the vast amount of data being produced. One of the problems related with this is the significant amount of missing interaction scores that cannot be scanned by the screening technologies or were filtered out from the datasets for technical reasons. This will significantly affect and bias downstream analysis. Therefore, there is an immediate need to impute those missing data more precisely. This study evaluates existing missing value imputation techniques on large-scale quantitative data matrices from synthetic genetic array (SGA) and epistatic miniarray profiling (E-MAP) screening technologies. Different existing methods that are usually applied for imputation purposes were evaluated against various conditions and performance accuracies. This best performing imputation approach, based on weighted correlation between nearest-neighbors, is now modified and can be used in any gene interaction data. Hereby, this study removed the limitation of a method already developed for this purpose and gives a more flexible, optimized, and best performing method. This method can now be effectively used in the pre-processing of gene interaction scores by researchers towards a genome-wide analysis such as identification of global functional networks, gene clustering, etc. for a more accurate and less biased results and biological interpretations.

gashawbk@yahoo.com