Identifying homeologues, orthologues and paralogues in the grasses

A. W. Schreiber\(^{(1)}\), U. Baumann\(^{(1)}\), E. Kalashyan\(^{(1)}\) and J. Schwerdt\(^{(1,2)}\)

\(^{(1)}\)Australian Centre for Plant Functional Genomics, University of Adelaide, Adelaide, Australia
\(^{(2)}\)South Australian Partnership for Advanced Computing, University of Adelaide, Adelaide, Australia

Genome-wide phylogenetic orthologue and paralogue identification across species is computationally prohibitively expensive, so approaches to this task often rely on a two-step approach by employing a clustering step based on best BLAST hits prior to a phylogenetic analysis within each resulting cluster\(^{(1)}\). On the other hand it is known that, even for fully sequenced genomes, best BLAST hits often miss closest phylogenetic neighbours\(^{(2)}\). This can give rise to incomplete phylogenetic reconstructions, a problem that gets compounded if the genomes being considered are not yet completely sequenced or if one needs to additionally identify homeologues because one or more of the included species are polyploid. This is the case for agronomically important grass species such as wheat and barley: neither of these two species is sequenced and wheat has a hexaploid genome. In order to carry out a comprehensive phylogenetic analysis of grasses such as these we have developed an alternate, efficient, graph-based clustering approach which includes more than just best BLAST hits. The method works because it still results in clearly separated clusters which can then be analyzed individually using phylogenetic methods. We present results and compare these to those obtained with other published methods.

References:
