

Ryegrass ASTRA: a web-based resource for *Lolium* EST analysis

G.C. Spangenberg^{1,2,3}, T. Sawbridge^{1,2}, E.K. Ong¹, C.G. Love^{1,2}, T.A. Erwin^{1,2}, E.G. Logan^{1,2} and D. Edwards^{1,2,3}
¹Primary Industries Research Victoria, Plant Biotechnology Centre, La Trobe University, Bundoora, Victoria 3086, Australia ²Victorian Bioinformatics Consortium, Australia ³Molecular Plant Breeding Cooperative Research Centre, Australia Email: german.spangenberg@dpi.vic.gov.au

Keywords: perennial ryegrass, expressed sequence tags (ESTs), gene annotation, bioinformatics ASTRA database, simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs)

Introduction Perennial ryegrass (*Lolium perenne* L.) is a major grass species of temperate pastoral agriculture.

Materials and methods A resource of 44,524 perennial ryegrass expressed sequence tags (ESTs) from single pass DNA sequencing of randomly selected clones from 29 cDNA libraries representing a range of plant organs and developmental stages was generated (Sawbridge *et al.*, 2003). Each of these sequences has been annotated by comparison to the GenBank and SwissProt public sequence databases. Automated intermediate Gene Ontology (GO) annotation has also been determined.

Results and conclusions Sequence annotation was determined for 80%, 53% and 50% of sequences using GenBank, SwissProt and GO annotation respectively. Only 17% of sequences were determined to have no comparative annotation and thus appear to be unique to ryegrass. EST redundancy was resolved through assembly with CAP3, identifying at total of 12,170 unigenes within this dataset. All sequences and annotation are maintained within an ASTRA format MySQL database, with web based access for text searching, BLAST sequence comparison and Gene Ontology hierarchical tree browsing. Each ryegrass sequence was mapped onto an EnSEMBL genome viewer for comparison with the complete genome sequence of rice and expressed sequences from related species.

In total 3,214 ryegrass simple sequence repeat (SSR) molecular markers have been identified using SSRPrimer and specific PCR amplification primers were designed. A total of 2,716 candidate single nucleotide polymorphism (SNP) molecular markers and 345 insertion deletions (InDels) molecular markers were also identified in 493 loci using AutoSNP. The integration of different analysis tools and the resulting data within a central resource allow users to identify novel candidate genes and associated molecular genetic markers that can be applied to functional genomics and germplasm enhancement in this agronomically important forage and turf grass.

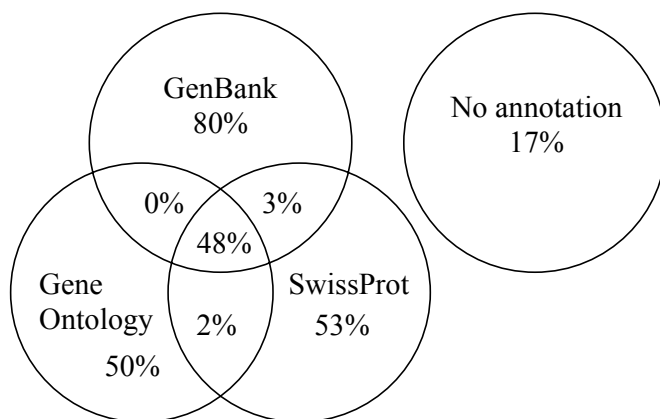


Figure 1 Annotation of ryegrass genes within the ASTRA database

Reference

Sawbridge, T., Ong, E.K., Binnion, C., Emmerling, M., McInnes, R., Meath, K., Nguyen, N., Nunan, K., O'Neill, M., O'Toole, F., Rhodes, C., Simmonds, J., Tian, P., Wearne, K., Webster, T., Winkworth, A. and Spangenberg, G. (2003). Generation and analysis of expressed sequence tags in perennial ryegrass (*Lolium perenne* L.). *Plant Science*, 165, 1089-1100.