

## Clover ASTRA: a web-based resource for *Trifolium* EST analysis

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

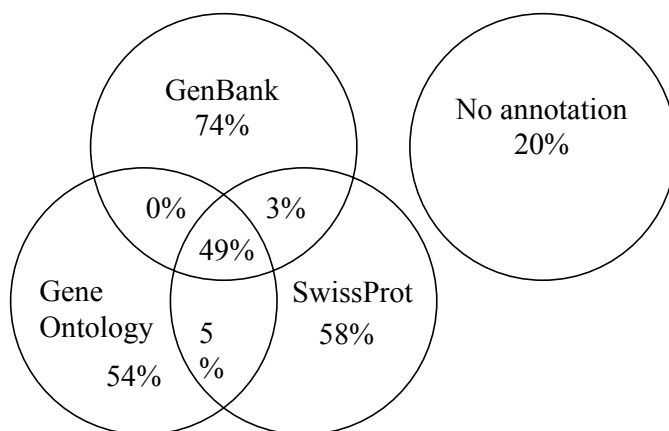
**Keywords:** white clover, expressed sequence tags (ESTs), gene annotation, bioinformatics ASTRA database, simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs)

**Introduction** White clover (*Trifolium repens* L.) is a major temperate forage legume.

**Materials and methods** A resource of 42,017 white clover expressed sequence tags (ESTs) from single pass DNA sequencing of randomly selected clones from 16 cDNA libraries representing a range of plant organs, developmental stages and environmental treatments 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 74%, 58% and 54% of sequences using GenBank, SwissProt and GO annotation respectively, with only 20% of sequences determined to have no comparative annotation and thus appear to be unique to white clover (Fig 1). EST redundancy was resolved through assembly with CAP3, identifying at total of 15,989 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 white clover sequence was mapped onto an EnSEMBL genome viewer for comparison with the complete genome sequence of *Arabidopsis thaliana* and expressed sequences from related legumes.

In total 5,407 white clover simple sequence repeat (SSR) molecular markers were identified using SSRPrimer and specific PCR amplification primers were designed. 18,517 candidate single nucleotide polymorphism (SNP) molecular markers and 1,706 insertion deletions (InDels) were also identified across 1,409 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 crop.



**Figure 1** Annotation of white clover genes within the ASTRA database

### Reference

Sawbridge, T., Ong, E.K., Binnion, C., Emmerling, M., Meath, K., Nunan, K., O'Neill, M., O'Toole, F., Simmonds, J., Wearne, K., Winkworth, A. and Spangenberg, G. (2003). Generation and analysis of expressed sequence tags in white clover (*Trifolium repens* L.). *Plant Science*, 165, 1077-1087.