

## Endophyte ASTRA: a web-based resource for *Neotyphodium* and *Epichloë* EST analysis

K. Shields<sup>1,2</sup>, M. Ramsperger<sup>1,2</sup>, S.A. Felitti<sup>1,2</sup>, C.G. Love<sup>1,3</sup>, T.A. Erwin<sup>1,3</sup>, D. Singh<sup>1,3</sup>, E.G. Logan<sup>1,3</sup>, D. Edwards<sup>1,2,3</sup> and G.C. Spangenberg<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>Molecular Plant Breeding Cooperative Research Centre, Australia <sup>3</sup>Victorian Bioinformatics Consortium, Australia Email: kate.shields@dpi.vic.gov.au

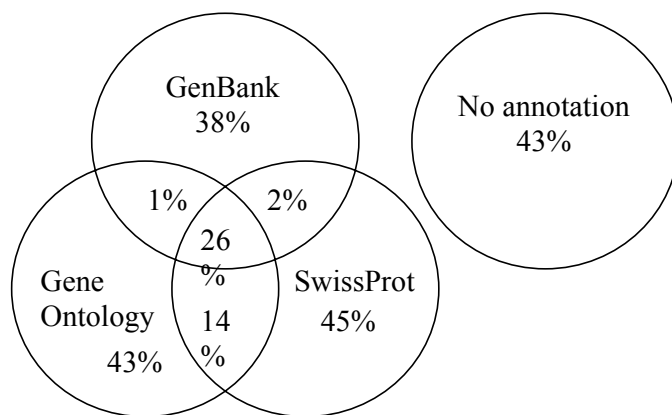
**Keywords:** grass endophytes, expressed sequence tags (ESTs), gene annotation, bioinformatics ASTRA database, simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs)

**Introduction** Large-scale gene discovery has led to the production of 13,964 expressed sequence tags (ESTs) collectively from the grass endophytes *Neotyphodium coenophialum*, *N. lolii* and *Epichloë festucae*.

**Materials and methods** A total of 9,783 ESTs were generated from fungal tissue grown *in vitro*, while 4,181 ESTs were generated from tissue grown *in planta*. Each of these sequences was annotated by comparison to the GenBank and SwissProt public sequence databases. Automated intermediate Gene Ontology (GO) annotation was also determined.

**Results and conclusions** Sequence annotation was determined for 38%, 45% and 43% of sequences using GenBank, SwissProt and GO annotation respectively, with 43% of sequences were determined to have no comparative annotation and thus appear to be unique to grass endophytes (Fig 1). EST redundancy was resolved through assembly with CAP3, identifying at total of 7,585 unigenes within this dataset, comprised of 5,846 singletons and 1,739 contigs. 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.

In total, 1,047 endophyte simple sequence repeat molecular markers (SSRs) have been identified using SSRPrimer and specific PCR amplification primers were designed. A total of 1,636 candidate single nucleotide polymorphism (SNP) molecular markers and 326 insertion deletions (InDels) were also identified across 300 loci using AutoSNP. The integration of different analysis tools and the resulting data within a central resource allow users to identify novel candidate genes for functional genomics and design of novel endophyte-grass associations.



**Figure 1** Annotation of endophyte genes within the ASTRA database