

Sequence Clustering of *Physcomitrella* ESTs and Comparison to Higher Plants

Martin Lomas¹ Celia Knight¹ David Westhead²

1 School Of Biology, University of Leeds, Leeds, LS2 9JT, UK.

2 School of Biochemistry and Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

The *Physcomitrella* EST Project (PEP) has been in place in Leeds since 1999. PEP has aimed to describe the transcriptome of moss *via* sequencing of ESTs obtained from tissue grown with different hormone treatments.

Bioinformatics techniques have been developed and applied to effect a computerised normalisation of *Physcomitrella* ESTs *via* clustering and consensus generation. Of the 14044 currently-sequenced ESTs, we have generated 8834 individual consensi.

Sequence comparison to the *Arabidopsis* and *Oryza* ORFeomes has been carried out using BLASTX has shown that 65% of these contigs have homologs in *Arabidopsis* at E values equal to or less than 0.001 and that 44% have homologues in *Oryza* at the same level. Analysis of the length of these matches has shown that over 70% of these have a homology length of over 200bp.

TBLASTN searches against the *Physcomitrella* consensi database have shown that 44% of the *Arabidopsis* ORFeome and 31% of the *Oryza* ORFeome are now represented in the PEP consensus sequence database.

Strategies are being developed and applied to select clones of interest for microarray expression studies. Results from these analyses are being made available at the PEP website: <http://www.moss.leeds.ac.uk> .