Invited Talk:

Annotating MADS-box Genes and KNOX Genes in the *Physcomitrella* Genome

Elizabeth I. Barker and Neil W. Ashton

Biology Department, University of Regina, Regina, Canada

Computational and experimental annotations make genomic sequence comprehensible and useful, predicting genes and providing information about their functions. The Joint Genome Institute's automatic annotation pipeline predicted genes at the scale of the entire *Physcomitrella* genome by *ab initio* and homology methods, using statistical analyses to locate features of genes and regulatory motifs and identifying sequences that were similar to known genes and available cDNAs. Gene models representing protein coding sequences, introns and, where possible, UTRs were constructed for known genes and previously unknown putative genes. Functions were predicted for putative gene products that were similar in sequence to proteins of known function. Gene prediction software is not infallible, however, and cannot provide complete information about genes and their functions. Manual annotation by the *Physcomitrella* community was, and continues to be, required to improve gene models and provide functional data. We annotated two families of transcription factors, the MADS-box gene family and the KNOX gene family. We found 26 MADS-box genes (including 2 pseudogenes), of which approximately half were previously unknown, and 6 KNOX genes, including one pseudogene and one other previously unknown gene, in the *Physcomitrella* genome. For the majority of the gene models, we were able to improve the exon-intron architectures and the sequences of the putative protein products. Where EST evidence was available, we added UTR sequence to models that lacked UTRs. References to the literature were added to our annotations. Using examples from the two transcription factor families, we will illustrate some of the *ab initio* and homology methods we used at the scale of individual genes to annotate *Physcomitrella*'s complements of MADS-box genes and KNOX genes.