

## Comparison of the *Physcomitrella* gametophyte transcriptome to the *Arabidopsis* Genome

Tomoaki Nishiyama<sup>1</sup>, Tomomichi Fujita<sup>1</sup>, Tadasu Shin-I<sup>2</sup>, Motoaki Seki<sup>3,4</sup>,  
Hiroyo Nishide<sup>1</sup>, Ikuo Uchiyama<sup>1</sup>, Asako Kamiya<sup>4</sup>, Piero Carninci<sup>3</sup>, Yoshihide Hayashizaki<sup>3</sup>,  
Kazuo Shinozaki<sup>3,4</sup>, Yuji Kohara<sup>2</sup>, and Mitsuyasu Hasebe<sup>1</sup>

- 1 National Institute for Basic Biology, Okazaki 444-8585, Japan.
- 2 National Institute of Genetics, Mishima, Japan.
- 3 RIKEN, Wako, Japan.
- 4 RIKEN Genomic Sciences Center, Yokohama, Japan.

The moss, *Physcomitrella patens*, diverged early in land plant evolution from the vascular plant lineage. To understand the change of genomic content during land plant evolution, genome-wide similarity of *Physcomitrella* and *Arabidopsis* was assessed by comparing expressed sequence tag (EST) data of *Physcomitrella* and sequence data of all the protein in *Arabidopsis*. We constructed full-length enriched cDNA libraries from, auxin-treated, cytokinin-treated, and non-treated gametophytes of *Physcomitrella*, and sequenced more than 40,000 clones from both ends. The ESTs are assembled with mRNA sequences of *Physcomitrella* deposited in GenBank, and 23,000 contigs were obtained. Contigs corresponding to the 5' and 3' end of the same clone were paired. Based on these pairs, our clones correspond to 13,000 independent mRNA. We constructed a database of *Physcomitrella* EST, containing the BLASTX search result of every contigs and clones. Currently, 15 Mb of *Physcomitrella* non-redundant EST can be searched with BLAST program in our database.

TBLASTN searches against the *Physcomitrella* database showed that 65% of *Arabidopsis* genes has a homologue in *Physcomitrella* with an *E* value less than or equal to 0.001. Conversely, all the moss contigs (23,000) were used as query in BLASTX searches against nr dataset. Of these, 10,000 showed the highest similarity to a vascular plant gene, and 12,000 found no hit with an *E* value  $\leq 0.001$ . The remaining 822 contigs had the highest similarity to a gene in an organism other than vascular plants, such as Metazoa and Bacteria, and included genes related to ion transport, DNA damage repair, which may reflect moss specific features. As gametophytes are extremely reduced in Angiosperms, some of the genes functioned in gametophytes may have lost, while most continued to function in sporophytes.