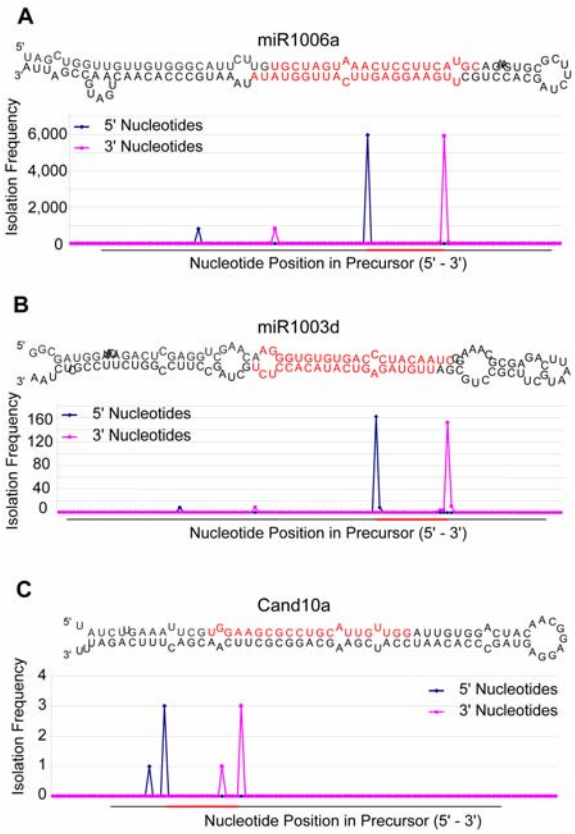
High-throughput sequencing of *Physcomitrella patens* small RNAs reveals many moss microRNAs and a conserved trigger for siRNA biogenesis.

Michael J. Axtell[1], Ramya Rajagopalan[1,2], and David Bartel[1,2,3]
[1]Whitehead Institute, [2]Massachusetts Institute of Technology, Biology Dept., and [3]Howard Hughes Medical Institute, 9 Cambridge Center, Cambridge, MA 02142-1479 USA

axtell@wi.mit.edu , PH: (+1) 617-258-8346, FAX (+1) 617-258-6768

MicroRNAs (miRNAs) and other endogenous silencing RNAs have important gene-regulatory roles in plants and animals. To understand better the scope and depth of RNA silencing pathways, we applied high-throughput pyrosequencing to the task of small RNA discovery and profiling. Sequencing of small RNA libraries from the moss *Physcomitrella patens* revealed more than 125,000 unique RNAs, most of which were sequenced only once. By comparing sequenced small RNAs to the whole-genome shotgun traces generated during the *P. patens* genome sequencing project we conclude that *P. patens* has more than 200 miRNA loci, which represent at least 89 miRNA families—a diversity that apparently exceeds that of higher plants. *P. patens* also expresses phased siRNAs, which resemble the trans-acting siRNAs important for leaf development in *Arabidopsis*. In both *P. patens* and *Arabidopsis*, phased siRNAs are generated from fragments that fall between miRNA complementary sites as well as fragments that fall between cleavage sites, indicating that dual complementary sites constitute a conserved trigger for siRNA biogenesis.



RNAs from representative *P. patens* microRNA loci.
(A) Small RNAs deriving from the miR1006a locus, one of 165 miRNA loci for which both the miRNA and the miRNA* were sequenced. The predicted RNA secondary structure of the miRNA hairpin is shown, with the most abundant small RNAs from each arm of the foldback highlighted in red. The isolation frequencies at each position of the foldback are plotted, with blue indicating the 5' terminus and pink indicating the 3' terminus of sequenced RNAs. The black line indicates the region of the locus corresponding to the predicted secondary structure; the red line indicates the region corresponding to the miRNA.
(B) Small RNAs deriving from the miR1003d locus, one of 36 loci for which the exact miRNA* was not sequenced, but other very closely related small RNAs corresponding to the star arm of the foldback were observed. Drawing conventions are as in (A).
(C) Small RNAs deriving from the cand10a locus, one of 17 miRNA candidate loci for which no sequences corresponding to the star arm of the foldback were observed.